

# Product-state Approximations to Quantum Ground States

Fernando G.S.L. Brandão\*

Aram W. Harrow<sup>†</sup>

October 2, 2013

## Abstract

The local Hamiltonian problem consists of estimating the ground-state energy (given by the minimum eigenvalue) of a local quantum Hamiltonian. It can be considered as a quantum generalization of constraint satisfaction problems (CSPs) and has a central role both in quantum many-body physics and quantum complexity theory. A key feature that distinguishes quantum Hamiltonians from classical CSPs is that the solutions may involve complicated entangled states. In this paper, we demonstrate several large classes of Hamiltonians for which product (i.e. unentangled) states can approximate the ground state energy to within a small extensive error.

First, we show the *existence* of a good product-state approximation for the ground-state energy of 2-local Hamiltonians with one or more of the following properties: (1) high degree, (2) small expansion, or (3) a ground state with sublinear entanglement with respect to some partition into small pieces. The approximation based on degree is a surprising difference between quantum Hamiltonians and classical CSPs, since in the classical setting, higher degree is usually associated with *harder* CSPs. The approximation based on low entanglement, in turn, was previously known only in the regime where the entanglement was close to zero.

Estimating the energy is NP-hard by the PCP theorem, but whether it is QMA-hard is an important open question in quantum complexity theory. A positive solution would represent a quantum analogue of the PCP theorem. Since the existence of a low-energy product state can be checked in NP, the result implies that any Hamiltonian used for a quantum PCP theorem should have: (1) constant degree, (2) constant expansion, (3) a “volume law” for entanglement with respect to any partition into small parts. The result also gives a no-go to a quantum version of the PCP theorem *in conjunction* with parallel repetition for quantum CSPs.

Second, we show that in several cases, good product-state approximations not only exist, but can be found in polynomial time: (1) 2-local Hamiltonians on any planar graph, solving an open problem of Bansal, Bravyi, and Terhal, (2) dense  $k$ -local Hamiltonians for any constant  $k$ , solving an open problem of Gharibian and Kempe, and (3) 2-local Hamiltonians on graphs with low threshold rank, via a quantum generalization of a recent result of Barak, Raghavendra and Steurer.

Our work introduces two new tools which may be of independent interest. First, we prove a new quantum version of the de Finetti theorem which does not require the usual assumption of symmetry. Second, we describe a way to analyze the application of the Lasserre/Parrilo SDP hierarchy to local quantum Hamiltonians.

\*Department of Computer Science, University College London, email: [fgslbrandao@gmail.com](mailto:fgslbrandao@gmail.com)

<sup>†</sup>Center for Theoretical Physics, Massachusetts Institute of Technology, email: [aram@mit.edu](mailto:aram@mit.edu)

# 1 Background

A quantum  $k$ -local Hamiltonian on  $n$  qudits is a  $d^n \times d^n$  Hermitian matrix  $H$  of the form

$$H = \frac{1}{l} \sum_{i=1}^l H_i, \quad (1)$$

where each term  $H_i$  acts non-trivially on at most  $k$  qudits<sup>1</sup> and  $\|H_i\| \leq 1$ . Local Hamiltonians are ubiquitous in physics, where interactions are almost always few-body. Of particular interest – e.g. in quantum many-body physics, quantum chemistry, and condensed matter physics – is to understand the low-energy properties of local Hamiltonians. These can be either the low-lying spectrum of the model, or properties – such as correlation functions – of the ground state of the Hamiltonian, defined as the eigenstate associated to the minimum eigenvalue. A benchmark problem, termed the local Hamiltonian problem, is to approximate the ground-state energy (i.e. the minimum eigenvalue) of the model. What is the computational complexity of this task?

In a seminal result Kitaev proved that the local Hamiltonian problem is complete for the quantum complexity class QMA, in the regime of estimating the energy to within polynomial accuracy in the number of particles of the model [50]. The class QMA is the quantum analogue of NP in which the proof consists of a quantum state and the verifier has a quantum computer to check its validity [77]. Since then there have been several developments showing that simpler classes of local Hamiltonians are still hard [49, 58, 70], culminating in the QMA-completeness of the problem for Hamiltonians on a line [6] even when all the local terms are the same and the only input is the length of the line expressed in unary [40]. In fact a new area has emerged around the question of understanding the computational complexity of local Hamiltonians, at the crossover of quantum complexity theory and condensed matter physics: quantum Hamiltonian complexity [59].

It is an important observation that the QMA-hardness of the local Hamiltonian problem implies, assuming  $\text{QMA} \neq \text{NP}$ , that one cannot find an efficient classical description of the ground state of local models in general. An efficient classical description is a representation of the quantum state in terms of polynomially many bits that would allow for the computation of local observables (e.g. the energy) in polynomial time. The QMA-hardness results therefore say that ground states of certain local Hamiltonians can be highly non-classical objects. In particular they can be highly entangled states. Indeed the local Hamiltonian problem can be seen as a quantum analogue of constraint satisfaction problems (CSPs), with the key distinguishing feature of quantum Hamiltonians being that the solutions may involve complicated entangled states.

A popular technique in physics to cope with this further level of complexity in the quantum case is to consider a product (i.e. unentangled) state as an *ansatz* to the true state. In quantum many-body physics this approximation is commonly referred to as *mean-field theory*, while in the context of quantum chemistry it is also called *Hartree-Fock method*. Conceptually, disregarding the entanglement in the ground state maps the quantum problem to a classical constraint satisfaction problem.

In this paper we give several new results concerning the usefulness of this product-state approximation for a large class of Hamiltonians. One motivation is to obtain a deeper understanding of when entanglement is relevant and to develop new polynomial-time algorithms for estimating

---

<sup>1</sup>See Section 1.2 for a precise definition.

the ground energy of interesting classes of Hamiltonians. A second motivation is related to a central open problem in quantum Hamiltonian complexity: The existence of a quantum version of the PCP theorem.

## 1.1 The quantum PCP conjecture

Let  $\mathcal{C}$  be a CSP and let  $\text{unsat}(\mathcal{C})$  be the fraction of the total number of local constraints that are not satisfied in the best possible assignment for  $\mathcal{C}$ . Then one formulation of the Cook-Levin theorem is that it is NP-hard to compute  $\text{unsat}(\mathcal{C})$  for an arbitrary CSP  $\mathcal{C}$ . However this leaves open the possibility that one might be able to approximate  $\text{unsat}(\mathcal{C})$  to within small error in reasonable time.

The objective of the theory of approximation algorithms is to find out in which cases one can obtain in polynomial time approximate solutions to problems that are hard to solve exactly. This is by now a well-developed theory [74]. The quantum case, in contrast, is largely unexplored, with a few notable exceptions [13, 39, 8, 43]. It is also interesting to understand what are the limitations for obtaining approximation algorithms; this is the goal of the theory of hardness of approximation, where one is interested in proving the computational hardness of obtaining even approximate solutions. A landmark result in this theory, on which almost all other hardness-of-approximation results are based, is the PCP (Probabilistically Checkable Proof) Theorem [10, 11, 29]. One possible formulation in terms of 2-CSPs<sup>2</sup> is the following:

**Theorem 1** (PCP Theorem [10, 11, 29]). *There is a constant  $\varepsilon_0 > 0$  such that it is NP-hard to determine whether for a given 2-CSP  $\mathcal{C}$ ,  $\text{unsat}(\mathcal{C}) = 0$  or  $\text{unsat}(\mathcal{C}) \geq \varepsilon_0$ .*

As well as developing a quantum theory of approximation algorithms, a theory of hardness of approximation for QMA would also be an interesting development. Inspired by the PCP theorem we might be tempted to speculate a quantum hardness-of-approximation result for the local Hamiltonian problem, which would correspond to a quantum analogue of the PCP theorem. In more detail, consider a local Hamiltonian  $H = \mathbb{E}_i H_i$  where  $\mathbb{E}_i$  refers to taking the expectation over  $i$  with respect to some distribution, and “local” means that no  $H_i$  acts on more than a constant number of qubits. The ground-state energy is defined by the smallest eigenvalue of  $H$ :

$$e_0(H) := \min\{\langle\psi|H|\psi\rangle : \langle\psi|\psi\rangle = 1\}. \quad (2)$$

The ground-state energy is the quantum analogue of the unsat value of CSPs. A central open problem in quantum Hamiltonian complexity is the validity of the following [59, 8, 43, 2, 1, 22, 61, 3]:

**Conjecture 2** (Quantum PCP Conjecture). *There is a constant  $\varepsilon_0 > 0$  such that it is QMA-hard to decide whether for a 2-local Hamiltonian  $H$ ,  $e_0(H) \leq 0$  or  $e_0(H) \geq \varepsilon_0$ .*

We remark that  $H$  is not necessarily positive semidefinite, so the case of  $e_0(H) \leq 0$  does not necessarily correspond to perfect completeness.

One could also consider the conjecture for  $k$ -local Hamiltonians. However Conjecture 2 is known to be equivalent to this more general version, even when restricted to 2-local Hamiltonians with sites formed by qubits [22].

---

<sup>2</sup>a  $k$ -CSP is a CSP in which each constraint only involves  $k$  variables.

There are many reasons why the quantum PCP conjecture is interesting (although a caveat to the following is that the quantum PCP conjecture, like the classical PCP theorem, is known not to hold in any constant number of spatial dimensions [13]).

1. Like the classical PCP theorem, a quantum PCP theorem would yield a hardness-of-approximation result for a QMA-hard problem at the error scale that is most natural. Moreover one could expect that, similarly to the PCP theorem, a quantum PCP theorem would lead to hardness-of-approximation results for other QMA-hard problems (e.g. see the survey [19]).
2. One of the central goals of theoretical, and especially condensed-matter, physics, is to explain how the properties of a quantum system, starting with its ground-state energy, emerge from its underlying Hamiltonian. A common method for doing so is to use insight and inspired guesswork to come up with an *ansatz*, which is an approximate ground state with a succinct classical description. If the quantum PCP conjecture were true, this would mean that there would exist families of Hamiltonians for which succinct, efficiently-verifiable, ansatzes do not exist, even if we accept constant extensive error. A related question is the NLTS (“no low-energy trivial states”) conjecture from [36], which can be equivalently stated as ruling out the specific family of ansatzes obtained by applying a constant-depth circuit to a fixed product state.
3. A quantum PCP conjecture would imply the QMA-hardness of approximating the energy of thermal states of local Hamiltonians at *fixed* (but sufficiently small) temperature. However on physical grounds one might expect that genuine quantum effects should be destroyed by the thermal fluctuations. A resolution of the quantum PCP conjecture would give new insight into this question.

See [3] for a recent review on the significance and status of the quantum PCP conjecture.

A way to *refute* the quantum PCP conjecture is to show that the estimation of the ground-state energy can be realized in NP<sup>3</sup>. In this work we follow this approach and give new approximation guarantees for the local Hamiltonian problem in the regime of extensive error.

## 1.2 Notation

Let  $\mathcal{D}(\mathcal{H})$  be the set of quantum states on  $\mathcal{H}$ , i.e. positive semidefinite matrices of unit trace acting on the vector space  $\mathcal{H}$ . A state on systems  $ABC$  is sometimes written  $\rho^{ABC}$  and we write  $\rho^A$  to denote the marginal state on the  $A$  system, i.e.  $\text{tr}_{BC}(\rho^{ABC})$ . For a pure state  $|\psi\rangle$ , we use  $\psi$  to denote the density matrix  $|\psi\rangle\langle\psi|$ .

We will measure entropy in “nats” so that the entropy of a density matrix  $\rho$  is defined as  $S(\rho) := -\text{tr}[\rho \ln(\rho)]$ . Entropic quantities are often written in terms of the subsystems; i.e. if  $\rho$  is a state on  $ABC$ , then  $S(ABC)_\rho := S(\rho^{ABC})$ ,  $S(A)_\rho := S(\rho^A)$ , etc. When the state is understood from context, we may suppress the subscript. Define the mutual information  $I(A : B) = S(A) + S(B) - S(AB)$  and the conditional mutual information  $I(A : B|C) = S(AC) + S(BC) - S(ABC) - S(C)$ . Additional information-theoretic quantities are defined in Section 3.2.

---

<sup>3</sup>Note it is extremely unlikely that this estimation can be done in P or even in BQP, since we know it is NP-hard by the PCP theorem.

Let  $\|X\|_1 := \text{tr} \sqrt{X^\dagger X}$  be the trace norm of  $X$ , and let  $\|X\|$  denote the operator norm, or largest singular value, of  $X$ .

Given a graph  $G = (V, E)$  and a subset of vertices  $S$ , we define the expansion of  $S$  as

$$\Phi_G(S) := \Pr_{(u,v) \in E} [v \notin S | u \in S]. \quad (3)$$

The expansion of the graph  $G$  is defined as

$$\Phi_G := \min_{S: |S| \leq |V|/2} \Phi_G(S). \quad (4)$$

Let  $H_d$  denote the set of Hermitian  $d \times d$  matrices. Define the set of  $k$ -local Hamiltonians to be

$$W_k = W_k^{(d,n)} = \text{Span}\{A_1 \otimes \cdots \otimes A_n : A_1, \dots, A_n \in H_d, \text{ at most } k \text{ of the } A_i \text{ are } \neq I_d\}. \quad (5)$$

In the language of quantum error-correction, the  $W_k$  are the operators of weight  $\leq k$ . For those familiar with Fourier analysis of Boolean functions, the diagonal matrices in  $W_k$  are precisely the functions of  $\{0, 1\}^n$  with all nonzero Fourier coefficients of weight  $\leq k$ .

Say that a state  $\sigma$  is globally separable if it can be written as  $\sum_z p_z \sigma_z$  for some  $\sigma_z$  of the form  $\sigma_z = \sigma_z^{Q_1} \otimes \cdots \otimes \sigma_z^{Q_n}$  and a probability distribution  $p_z$ . A state is separable with respect to a partition  $V_1 : \dots : V_m$  if it can be written in the form  $\sum_z p_z \sigma_z^{V_1} \otimes \cdots \otimes \sigma_z^{V_m}$ .

## 2 Results

Our main results concern Hamiltonians for which product states give good approximations of the ground-state energy. In Section 2.1, we demonstrate a large class of Hamiltonians with this property, which puts the corresponding ground-state estimation problem in NP.

In Section 2.2, we give three cases in which such a product-state approximation can be found efficiently: planar Hamiltonians in Section 2.2.1, dense Hamiltonians in Section 2.2.2 and Hamiltonians on graphs with low threshold rank (defined below) in Section 2.2.3.

Each of these results, except for the case of low threshold rank (which follows the analysis of the Lasserre hierarchy in [14]) uses an information-theoretic approach based on the chain rule of conditional mutual information, and is inspired by a method due to Raghavandra and Tan [63] originally developed to prove fast convergence of the Lasserre/Parrilo/sum-of-squares (SOS) hierarchy for certain CSPs. Similar techniques are also used to prove several new quantum de Finetti theorems in our companion paper [20].

Indeed, de Finetti theorems are another sort of product-state approximation and our proof approach can be seen as a de Finetti theorem. We explore this in Section 2.3 by stating a new de Finetti theorem which is the first version of the de Finetti theorem to *not* require any symmetry assumptions.

The low-threshold-rank result in Section 2.2.3 is based on a variant of the SOS hierarchy that has been widely used in quantum chemistry [33, 56, 82, 54, 15, 16], but until now only as a heuristic algorithm.

## 2.1 Approximation Guarantees in NP

Our first and main result shows that one can obtain a good approximation to the ground-state energy (or indeed the energy of any state) by a product state for a large class of 2-local Hamiltonians. To every 2-local Hamiltonian we can associate a constraint graph  $G = (V, E)$  with  $V$  the set of vertices associated to the sites of the Hamiltonian, and  $E$  the set of edges associated to the interaction terms of the Hamiltonian. The basic result is that if  $G$  is a  $D$ -regular graph then product states can approximate the ground energy with an error that is a power of  $1/D$ , yielding a good approximation for sufficiently large constant degree  $D$ .

The rest of this section is organized as follows. The basic version of our result described above is given in Section 2.1.1. Then we proceed to two different generalizations. In Section 2.1.2, we consider  $D$ -regular graphs in which groups of vertices have been clustered together. In this case, our approximation is further improved if the clusters are less than maximally expanding, or are less than maximally entangled. Then in Section 2.1.3, we give a version of the theorem for variable-degree weighted graphs in which  $1/D$  is replaced by the collision probability (probability that a two-step random walk starting from a random vertex will return to its starting point). Then in Section 2.1.4 we discuss the significance of the results for the quantum PCP conjecture and connect them to previous work. An overview of the proof is given in Section 2.1.5. Finally in Section 2.1.6 we discuss some additional variants of main theorem.

### 2.1.1 Regular graphs

**Theorem 3** (Basic product-state approximation). *Let  $G = (V, E)$  be a  $D$ -regular graph with  $n = |V|$ . Let  $\rho^{Q_1 \dots Q_n}$  be an  $n$ -qudit state. Then there exists a globally separable state  $\sigma$  such that*

$$\mathbb{E}_{(i,j) \in E} \|\rho^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 \leq 12 \left( \frac{d^2 \ln(d)}{D} \right)^{1/3}. \quad (6)$$

Theorem 3 applies to any state  $\rho$ . If  $\rho$  is the ground state of a Hamiltonian then we obtain:

**Corollary 4.** *Let  $G = (V, E)$  be a  $D$ -regular graph, and let  $H$  be a Hamiltonian on  $n = |V|$   $d$ -dimensional particles given by*

$$H = \mathbb{E}_{(i,j) \in E} H_{i,j} = \frac{2}{nD} \sum_{(i,j) \in E} H_{i,j},$$

where  $H_{i,j}$  acts only on qudits  $i, j$  and satisfies  $\|H_{i,j}\| \leq 1$ . Then there exists a state  $|\varphi\rangle$  such that  $|\varphi\rangle = |\varphi\rangle^{Q_1} \otimes \dots \otimes |\varphi\rangle^{Q_n}$  and

$$\text{tr}(H\varphi) \leq e_0(H) + 12 \left( \frac{d^2 \ln(d)}{D} \right)^{1/3} \quad (7)$$

**Corollary 5.** *Let  $G, H$  be as in Corollary 4. Consider the promise problem of estimating whether  $e_0(H) \leq \alpha$  or  $e_0(H) \geq \beta$ . If  $\beta - \alpha \geq 12 \left( \frac{d^2 \ln(d)}{D} \right)^{1/3} + \delta$  then there exists witnesses of length  $O(nd \log(n/\delta))$  for the  $e_0(H) \leq \alpha$  case. The verifier requires time  $nDd^4 \text{poly}(\log(1/\delta))$ .*

### 2.1.2 Clustered approximation for regular graphs

The first refinement of Theorem 3 we consider is when the graph admits a natural way to group the vertices into clusters. In case *either* the clusters are lightly expanding or are lightly entangled, we can improve on the bound of Theorem 3.

**Theorem 6** (Clustered product-state approximation for regular graphs). *Let  $G = (V, E)$  be a  $D$ -regular graph, with  $n = |V|$ ,  $m$  a positive integer dividing  $n$  and  $\{V_1, \dots, V_{n/m}\}$  a partition of  $V$  into sets each of size  $m$ . Define  $\bar{\Phi}_G := \mathbb{E}_{i \in [n/m]} \Phi_G(V_i)$  to be the average expansion of the blocks.*

*Let  $\rho$  be a state on  $(\mathbb{C}^d)^{\otimes n}$  and define  $\bar{I} := \mathbb{E}_{i \in [n/m]} I(V_i : V_i^c)_\rho$  to be the average mutual information of  $V_i$  with its complement for state  $\rho$ .*

*Then there exists a state  $\sigma$  such that*

1.  $\sigma$  is separable with respect to the partition  $(V_1, \dots, V_{n/m})$ ; i.e.  $\sigma = \sum_z p_z \sigma_z$  for some states  $\sigma_z$  of the form  $\sigma_z = \sigma_z^{V_1} \otimes \dots \otimes \sigma_z^{V_{n/m}}$ ; and

2.

$$\mathbb{E}_{(i,j) \in E} \|\rho^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 \leq 9 \left( \frac{d^2 \bar{\Phi}_G}{D} \frac{\bar{I}}{m} \right)^{1/6} + \frac{1}{m} + \frac{m}{n} \quad (8)$$

Along the lines of corollaries 4 and 5, we can apply Theorem 6 to the ground state of a local Hamiltonian.

**Corollary 7.** *Let  $G$  and  $V_1, \dots, V_{n/m}$  be as in Theorem 6, and let  $H$  be a Hamiltonian on  $n = |V|$   $d$ -dimensional particles given by*

$$H = \mathbb{E}_{(i,j) \in E} H_{i,j} = \frac{2}{nD} \sum_{(i,j) \in E} H_{i,j},$$

*where  $H_{i,j}$  acts only on qudits  $i, j$  and satisfies  $\|H_{i,j}\| \leq 1$ . Let  $\psi_0 = |\psi_0\rangle\langle\psi_0|$  be a ground state of  $H$ ; i.e.  $e_0(H) = \langle\psi_0|H|\psi_0\rangle$ . Then there exists  $|\varphi\rangle := |\varphi_1\rangle^{V_1} \otimes \dots \otimes |\varphi_{n/m}\rangle^{V_{n/m}}$  such that*

$$\text{tr}(H\varphi) \leq e_0(H) + 9 \left( \frac{d^2 \bar{\Phi}_G}{D} \frac{\mathbb{E}_i S(V_i) \psi_0}{m} \right)^{1/6} + \frac{1}{m} + \frac{m}{n}. \quad (9)$$

*There is a classical string of length  $O(n2^m \log(1/\delta))$  describing  $|\varphi\rangle$  to accuracy  $\delta$ .*

This description can be verified in time polynomial in the witness length. Thus, the resulting ground-state estimation problem is in NP when  $m = O(\log n)$  (or more generally has a subexponential-size witness when  $m = o(n)$ ).

### 2.1.3 Non-regular graphs:

In Theorems 3 and 6 our requirement that  $G$  be a regular graph was somewhat artificial. Indeed, we can also consider graphs with weighted edges. More generally, define

$$H := \sum_{(i,j) \sim G} H_{i,j} := \sum_{i,j} G_{i,j} H_{i,j}, \quad (10)$$

where each  $\|H_{i,j}\| \leq 1$  and  $G$  is a matrix satisfying  $G_{i,j} \geq 0$  and  $\sum_{i,j} G_{i,j} = 1$ . This is “PCP-style” weighting with the triangle inequality naturally yielding the bound  $\|H\| \leq 1$ , and with the quantum PCP conjecture concerning the hardness of an  $\epsilon$ -approximation to  $e_0(H)$  for sufficiently small constant  $\epsilon$ .

We can interpret  $G$  as a graph as follows. Without loss of generality we can assume that  $G = G^T$ . Define the vector  $\pi$  by  $\pi_j := \sum_i G_{i,j}$ , and observe that  $\pi$  is a probability distribution. Define  $A_{i,j} = G_{i,j}/\pi_j$ . Since  $\sum_i A_{i,j} = 1$  for any  $j$ ,  $A$  is a stochastic matrix. Further observe that  $A\pi = \pi$ . Thus we can interpret  $A$  as a random walk with stationary distribution  $\pi$ , and we can interpret  $G$  as the distribution on edges obtained by sampling a vertex according to  $\pi$  and then taking one step of the walk according to  $A$ . Because  $G = G^T$ , the random walk is reversible<sup>4</sup>.

In this case we can prove an analogue of Theorems 3 and 6 for weighted graphs, but without the ability to define clusters. Thus, our guarantee will not be in terms of expansion or entanglement, but only a quantity analogous to average degree.

**Theorem 8.** *Let  $G$  be a symmetric matrix with nonnegative entries summing to one and let  $A_{i,j} := G_{i,j}/\sum_{i'} G_{i',j}$ . Then for any  $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$  there exists a globally separable state  $\sigma$  such that*

$$\mathbb{E}_{(i,j) \sim G} \|\rho^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 \leq 14 \left( d^4 \ln(d) \operatorname{tr}[A^2] \|\pi\|_2^2 \right)^{1/8} + \|\pi\|_2^2 \quad (11)$$

*If further we have a Hamiltonian  $H$  given by Eq. (10), then there exists a state  $|\varphi\rangle = |\varphi_1\rangle \otimes \cdots \otimes |\varphi_n\rangle$  such that*

$$\operatorname{tr}(H\varphi) \leq e_0(H) + 14 \left( d^4 \ln(d) \operatorname{tr}[A^2] \|\pi\|_2^2 \right)^{1/8} + \|\pi\|_2^2 \quad (12)$$

The proof is in Section 5.2.

Degree here is replaced by  $\|\pi\|_2^2 \operatorname{tr}(A^2)$  and indeed it can be seen as the harmonic mean of the degree. Ideally the bound would be unchanged by trivial changes such as adding degree-0 vertices. One possible such strengthening of Theorem 8 would be to replace the average degree with the harmonic mean weighted by the measure  $\pi$ ; i.e.  $1/\sum_{i,j} \pi_i A_{i,j}^2$ .

**Conjecture 9.** *The  $\operatorname{tr}[A^2] \|\pi\|_2^2$  term in Theorem 8 can be replaced by some power of  $\sum_{i,j} \pi_i A_{i,j}^2$ .*

## 2.1.4 Discussion of the Results and Implications for the Quantum PCP Conjecture

Let us discuss the significance of Theorems 8, 6, and 3. Theorem 6 gives a non-trivial approximation in terms of three parameters: the average expansion  $\bar{\Phi}_G$  of the regions  $V_i$ , the degree  $D$  of the graph, and the average entanglement  $\bar{I}$  of the regions  $V_i$  with their complement. Below we discuss the relevance of each of them.

**Average Expansion:** This is the least surprising term, since one can always delete the edges between different blocks and incur an error that is upper bounded by the average expansion.

**Degree of the Graph:** We believe this is a more surprising approximation parameter. Indeed an easy corollary of the PCP theorem [10, 11, 29] is the following (see section 6 for a proof):

---

<sup>4</sup>Despite this rather specific interpretation, our construction is of course fully general. Indeed, suppose that  $\tilde{H} = \sum_{i,j} \tilde{H}_{i,j}$ . Let  $\Gamma = \sum_{i,j} \|\tilde{H}_{i,j}\|$ . Then we can define  $G_{i,j} = \|\tilde{H}_{i,j}\|/\Gamma$ ,  $H_{i,j} = \frac{\tilde{H}_{i,j}}{\|\tilde{H}_{i,j}\|}$  and we have that  $G$  is of the appropriate form and  $H := \tilde{H}/\Gamma = \sum_{i,j} G_{i,j} H_{i,j}$ .



**Proposition 10.** *For any constants  $c, \alpha, \beta > 0$ , it is NP-hard to determine whether a 2-CSP  $\mathcal{C}$ , with alphabet  $\Sigma$  on a constraint graph of degree  $D$ , is such that  $\text{unsat}(\mathcal{C}) = 0$  or  $\text{unsat}(\mathcal{C}) \geq c|\Sigma|^\alpha/D^\beta$ .*

Assuming  $\text{QMA} \not\subseteq \text{NP}$ , Theorem 6 shows the analogous result for the quantum case – namely that it is QMA-hard to achieve a  $cd^\alpha/D^\beta$  approximation for the ground-state energy of 2-local Hamiltonians – is *not* true. This suggests a route for a disproof of the quantum PCP conjecture by trying to find a procedure for increasing the degree of the constraint graph without decreasing the ground-state energy of the model (by possibly increasing the dimension of the particles as well).

Classically this can be achieved by parallel repetition of the CSP [64] or by gap amplification [29], but it is unclear whether there is an analogous construction in the quantum case. We note that such degree-increasing maps have been pursued in the quantum case with an eye toward proving the quantum PCP conjecture [2] whereas our results suggest that they may be useful in *disproving* it. Indeed a consequence of the result is that a quantization of the gap amplification part of Dinur’s proof of the PCP theorem [29] to the quantum setting would disprove the quantum PCP conjecture. Therefore we conclude that we cannot quantize Dinur’s approach without introducing substantial new ideas.

Parallel repetition gives a mapping  $\mathcal{P}_t$  from 2-CSP to 2-CSP, for every integer  $t$ , such that  $\deg(\mathcal{P}_t(\mathcal{C})) = \deg(\mathcal{C})^t$ ,  $|\Sigma_{\mathcal{P}_t(\mathcal{C})}| = |\Sigma_{\mathcal{C}}|^t$ <sup>5</sup>,  $\text{unsat}(\mathcal{P}_t(\mathcal{C})) \geq \text{unsat}(\mathcal{C})$ , and  $\text{unsat}(\mathcal{P}_t(\mathcal{C})) = 0$  if  $\text{unsat}(\mathcal{C}) = 0$  (see section 6 for more details). The next corollary shows that a similar notion of parallel repetition for quantum constraint satisfaction problems would disprove the quantum PCP conjecture.

**Corollary 11.** *Suppose that for every  $t \geq 1$  there is a mapping  $\mathcal{P}_t$  from 2-local Hamiltonians on  $d$ -dimensional particles to 2-local Hamiltonians on  $n_t$   $d_t$ -dimensional particles such that:*

- (i)  $\mathcal{P}_t$  can be computed deterministically in polynomial time.
- (ii)  $\mathcal{P}_t(H) = \mathbb{E}_{(i,j) \sim G'} H'_{i,j}$  with  $G'$  a probability distribution and  $\|H'_{i,j}\| \leq 1$ .
- (iii)  $\deg(\mathcal{P}_t(H)) \geq \deg(H)^t$ ,
- (iv)  $n_t \leq n^{O(t)}$ ,
- (v)  $d_t = d^t$ ,
- (vi)  $e_0(\mathcal{P}_t(H)) \geq e_0(H)$ , if  $e_0(H) > 0$ ,
- (vii)  $e_0(\mathcal{P}_t(H)) \leq e_0(H)$ , if  $e_0(H) \leq 0$ .

*Then the quantum PCP conjecture is false.*

The proof is in Section 6.

*Entanglement Monogamy and Mean Field:* The approximation in terms of the degree can be interpreted as an instance of monogamy of entanglement [73], a feature of quantum correlations that says a system cannot be highly entangled with a large number of other systems. If the degree is

---

<sup>5</sup> $\Sigma_{\mathcal{P}_t(\mathcal{C})}$  and  $\Sigma_{\mathcal{C}}$  are the alphabets of  $\mathcal{P}_t(\mathcal{C})$  and  $\mathcal{C}$ , respectively.

high, it means that every particle is connected to many others by an interaction term. But since it cannot be highly entangled with many of them, we might expect that a product state gives a reasonable approximation to the energy of most such interactions. Indeed based on this idea it is folklore in condensed matter physics that mean field (where one uses a product state as an ansatz to the ground state) becomes exact once the dimension of the interaction lattice increases. For rotationally symmetric models this can be rigorously established [52] using the standard quantum de Finetti Theorem [26]. The approximation we obtain here in terms of the degree of the graph gives a rigorous justification in the general case.

*Highly Expanding Graphs:* Finally the result also gives an approximation for *highly expanding* graphs. Using the bound  $\Phi_G \leq 1/2 - \Omega(\deg^{-1/2})$  (see e.g. Ref. [45]), Theorem 6 gives the existence of states  $|\varphi_1\rangle, \dots, |\varphi_n\rangle$  such that

$$\langle \varphi_1, \dots, \varphi_n | H | \varphi_1, \dots, \varphi_n \rangle \leq e_0(H) + \Omega \left[ \left( \frac{1}{2} - \Phi_G \right)^{1/3} \right]. \quad (13)$$

Therefore very good expanders, with  $\Phi_G$  approaching the maximum value  $1/2$ , are not candidates for QMA-hard instances. See also Section 2.2.3 for a similar setting in which not only do product states provide a good approximation, but a good choice of product state can be efficiently found.

A recent and independent result of Aharonov and Eldar [5] shows that the ground-state energy of commuting Hamiltonians on hypergraphs with almost perfect small-set expansion can be approximated in NP. This is weaker than our result in the sense of its restriction to commuting Hamiltonians, and its requirement of high small-set expansion (which is similar to, but strictly more demanding than our requirement of high degree), but it is stronger than our result in its applicability to  $k$ -local Hamiltonians with  $k > 2$ . To apply our proof techniques to  $k$ -local Hamiltonians for  $k > 2$  and obtain error  $\epsilon$ , we would need there to be  $n^{k-1}/\epsilon^{O(1)}$  terms, so only for  $k = 2$  could this be interpreted as “large constant degree.” Another possible way to obtain the results of [5] from ours would be to reduce a  $k$ -local Hamiltonian to a 2-local Hamiltonian using the perturbation-theory gadgets of [22]. Unfortunately the resulting 2-local Hamiltonian has a constant fraction of vertices with only constant degree, which means our results do not yield non-trivial bounds here. Indeed the gadgets of [22] create a ground state with significant entanglement, so any classical approximation of it would need to go beyond using product states.

**Average Entanglement:** Given a bipartite pure state  $|\psi\rangle^{AB}$  a canonical way to quantify the entanglement of  $A$  and  $B$  is by the von Neumann entropy of the reduced density matrix  $\psi^A$  (or  $\psi^B$ ), given by  $S(A)^\psi := -\text{tr}(\psi^A \ln \psi^A)$  [46]. Theorem 6 shows that we can obtain an ansatz of size  $n2^{O(m)}$  that approximates well the ground energy whenever the ground state satisfies a *subvolume* law for entanglement, i.e. whenever we can find a partition of the sites into regions  $V_i$  of size  $m$  such that in the ground state their entanglement (with the complementary region) is sublinear, on average, in the number of particles of the regions.

Although the dependence of our result on entanglement is intuitive, it is not trivial. One can trivially say (from Pinsker’s inequality) that if  $\mathbb{E}_i S(V_i)_\psi \leq \epsilon$  then there will exist a product state with energy within  $O(\sqrt{\epsilon})$  of that of  $\psi$ . However, we are very far from this regime, and obtain effective approximations whenever the entanglement is merely not within a constant factor of its maximum value.

There is a general intuition that the amount of entanglement in the ground state of a Hamil-

tonian should be connected to its computational complexity (see e.g. [32]). For one-dimensional models this can be made precise: An area law for entanglement in the ground state, meaning that the entanglement<sup>6</sup> of an arbitrary connected region is a constant, implies there is an efficient classical description for estimating the ground-state energy to within inverse polynomial accuracy [76, 75, 71] (in terms of a matrix-product state of polynomial bond dimension [34]).

Here we find a similar situation, but with notable differences. On one hand the result is weaker in that it only implies an efficient classical description of the ground state for estimating the ground-state energy; on the other hand the result is much more general, being true for 2-local Hamiltonians on general interaction graphs and only requiring a subvolume law instead of an area law. This new relation of entanglement and computational hardness gives a motivation for studying the entanglement scaling in ground states of Hamiltonians on more general graphs than finite dimensional ones.

Another interesting choice for the state  $\rho$  in Theorem 6 is the thermal state of the Hamiltonian at temperature  $T$ :  $\rho_T := \exp(-H/T) / \text{tr}(\exp(-H/T))$ . In this case, it is more convenient to normalize the Hamiltonian to have an extensive amount of energy; i.e.  $H = \sum_{i \in V} \mathbb{E}_{j \sim \Gamma(i)} H_{ij}$ , where  $\Gamma(i)$  is the set of neighbors of  $i$  and again each  $\|H_{i,j}\| \leq 1$ . Thus  $\|H\| \leq n$ . Now using the bound<sup>7</sup>  $\text{tr}(H\rho_T) \leq e_0(H) + \varepsilon$  for  $T = O(\varepsilon / \ln(d))$  we find there is a state  $|\varphi\rangle = |\varphi_1\rangle \otimes \dots \otimes |\varphi_{n/m}\rangle$  satisfying

$$\frac{1}{n} \text{tr}(H\varphi) \leq \frac{1}{n} e_0(H) + \varepsilon + 9 \left( \frac{d^2 \bar{\Phi}_G}{D} \frac{\mathbb{E}_i I(V_i : V_i^c)_{\rho_T}}{m} \right)^{1/6}. \quad (14)$$

Therefore Hamiltonians for which the thermal state satisfies a subvolume law for the mutual information for constant, but arbitrary small, temperatures are also not candidates for QMA-hard instances. We find this observation interesting since it is much easier to prove area laws for thermal states than for ground states [79]. We note however that the known area law for thermal states from Ref. [79] is not strong enough to give a non-trivial approximation for arbitrary 2-local Hamiltonians.

The approximation in terms of the average entanglement means that in order for the quantum PCP conjecture to be true we must find a family of Hamiltonians on qubits whose ground states satisfy a strict volume law for the entanglement of every region of size  $n^{o(1)}$  (on average over the choice of the region for every partition of the sites into  $n^{o(1)}$ -sized regions). Such examples can be constructed: For example, take a 3-regular expander and place three qubits at each vertex. Using 2-local interactions we can ensure that the ground state is the tensor product of an EPR pair along each edge. The expansion condition then ensures that this state satisfies a volume law. On the other hand such states are easy to give witnesses for using the techniques of [23, 43]. To the best of our knowledge there is no known example of a Hamiltonian that both satisfies a strict volume law and cannot be described by a witness from [43].

### 2.1.5 Proof sketch

The proofs of Theorems 3, 6 and 8 are based on information-theoretic methods. These methods were introduced by Raghavendra and Tan [63] (see also [14]) and we adapt them to our purposes.

<sup>6</sup>In this case one must quantify entanglement not by the von Neumann entropy, but by the smooth max-entropy [71].

<sup>7</sup>This bound can be derived from the fact that the Gibbs state  $\rho_T$  minimizes the free energy  $F(\rho) := \text{tr}(H\rho) - TS(\rho)$ .

(The motivation of [63, 14] was to analyze the Lasserre hierarchy for certain CSPs, and we explore a quantum version of the Lasserre hierarchy in Section 2.2.3.)

The key idea is that given a quantum state of  $n$  subsystems  $Q_1, \dots, Q_n$ , by measuring a small number  $k$  of the subsystems chosen at random, one can show that *on average* the remaining pairs of subsystems will be close to a product state up to small error. (Speaking about the average pair is crucial, as one can see by imagining  $n/2$  Bell pairs.) This statement is a consequence of the *chain rule* of the mutual information. As we show, the error scales as the quotient of the average mutual information of  $V_i$  with its complementary region and  $k$ , which we choose as  $k = \delta n$  for small  $\delta > 0$ . Then in a next step we show how an average pair of particles being close to product implies that an average pair of particles *connected* by the constraint graph of the Hamiltonian are close to product (increasing the error in the process); it is here that the degree of the graph appears as an approximation factor. So far, our strategy generally follows the iterative conditioning approach of [14], which applied to classical random variables. To bridge the gap to quantum states, we will need to use an informationally-complete measurement, which simply means a POVM from whose outcomes the original state of a system can be reconstructed.

For the product-state approximation, we choose the tensor product of all single-particle reduced states of  $\tilde{\rho}$ , with the state  $\tilde{\rho}$  being the postselected state obtained by measuring  $t$  subsystems of the original state  $\rho$  (which we typically take to be the ground state of the Hamiltonian). By the ideas outlined above, we can show that this state has energy no bigger than the R.H.S. of Eq. (8).

As a final ingredient in the clustered case (Theorem 6), before applying the argument above we fuse some of the original particles of the model into blocks  $V_1, \dots, V_{n/m}$  which we treat as individual particles. This creates new difficulties (such as the fact that the degree of the effective graph on the blocks is not the same as the original degree) that are handled by applying the chain rule of mutual information in two stages: first to the blocks of sites and then to the sites in each block.

### 2.1.6 Variants of the main theorems

In this section we sketch a few easy-to-prove variants of our main theorems.

Another way to state our main theorems is that if we measure a random small subset of qudits and condition on the outcomes, then the remaining qudits will be approximately product across a random edge of the graph. See Eq. (56) for the clustered version of this claim.

This version is also meaningful for classical probability distributions, in which case we obtain a variant of Lemma 4.5 of [63]. The result, roughly, is that conditioning on a random  $\delta n$  variables leaves *all* of the random variables approximately independent across a random edge. Now there is no problem from measurement disturbance, although conditioning may have its own costs.

All of our results make use of state tomography because we wish to consider arbitrary 2-local interactions. This is what leads to the  $\text{poly}(d)$  terms in the error. However, suppose the individual Hamiltonian terms  $H_{i,j}$  are of the form  $\sum_{x,y} \lambda_{x,y}^{i,j} A_x \otimes A_y$  where each  $A_x \geq 0$ ,  $\sum_x A_x = I$  and  $|\lambda_{x,y}^{i,j}| \leq 1$ . Then we can replace the generic state tomography procedure with the measurement  $\{A_x\}$ . This is essentially the approach used in [20] to prove de Finetti theorems for local measurements. The result is to remove the  $\text{poly}(d)$  terms in our error bounds.

The above claims are straightforward to prove, although we will not do so in this paper. We

also state one conjecture that we cannot prove. It is natural to conjecture a common generalization of Theorem 6 and Theorem 8 that yields a meaningful bound for clustered weighted graphs. We see no barrier to proving this apart from a lot of hairy algebra.

## 2.2 Efficient algorithms for finding product-state approximations

When our task is only to find a product state minimizing the energy of a  $k$ -local Hamiltonian, then essentially our problem becomes a classical  $k$ -CSP, albeit with an alphabet given by the set of all unit vectors in  $\mathbb{C}^d$ . However, we find that the problem is easy to approximate in many of the same cases that classical CSPs are easy to approximate, such as planar graphs, dense hypergraphs and graphs with low threshold rank.

### 2.2.1 Polynomial-time Approximation Scheme for Instances on a Planar Graph

It is known that there are polynomial-time approximating schemes (PTAS) for a large class of combinatorial problems (including many important NP-hard problems) on planar graphs [12]. Bansal, Bravyi, and Terhal generalized this result to obtain a PTAS for the local Hamiltonian problem on planar 2-local Hamiltonians with a bounded degree constraint graph [13]. They left as an open problem the case of unbounded degree. Combining their result with the techniques used in Section 2.1 we show:

**Theorem 12.** *For every  $\varepsilon > 0$  there is a polynomial-time randomized algorithm such that, given a 2-local Hamiltonian  $H$  on a planar graph, computes a number  $x$  such that  $e_0(H) \leq x \leq e_0(H) + \varepsilon$ .*

The idea for obtaining Theorem 12 is to use the same information-theoretic techniques behind Theorem 3 to replace the ground state by a state which is product on all high-degree vertices and the original ground state on the low-degree part, only incurring a small error in the energy. Then using the methods of [12, 13] we can also replace the state on the low-degree vertices by a tensor product of states each on a constant number of sites. At this point we have a constraint satisfaction problem over a planar graph and we can employ e.g. [12, 13] to approximate the optimal solution in polynomial time.

### 2.2.2 Polynomial-time Approximation Scheme for Dense Instances

We say a  $k$ -local Hamiltonian  $H = \frac{1}{m} \sum_{j=1}^m H_j$  is dense if  $m = \Theta(n^k)$ . Several works have established polynomial-time approximation schemes (PTAS) for dense CSPs [27, 35, 37, 9, 7, 17]. In Ref. [39] Gharibian and Kempe generalized these results to the quantum case showing that one also has a PTAS for estimating the energy of dense quantum Hamiltonians over product state assignments. They also asked whether one could improve the result to obtain a PTAS to the ground-state energy [39]. We answer this question in the affirmative:

**Theorem 13.** *For every  $\varepsilon > 0$  there is a polynomial-time algorithm such that, given a dense  $k$ -local Hamiltonian  $H$ , computes a number  $x$  such that  $e_0(H) \leq x \leq e_0(H) + \varepsilon$ .*

The idea behind the proof of Corollary 13 is to show that a product state assignment gives a good approximation to the ground-state energy of dense local Hamiltonians. The statement then

follows from the result of Ref. [39]. Theorem 3 indeed shows that product states give a good approximation in the case of 2-local Hamiltonians. The necessary work here is then to generalize the argument to  $k$ -local Hamiltonians for arbitrary  $k$ .

One approach might be to employ the perturbation-theory gadgets of Bravyi, DiVincenzo, Loss, and Terhal [22]. However, this approach ruins the density by creating a constant fraction of constant-degree vertices.

Instead we will address the case of  $k$ -local Hamiltonians directly. In the process, we will flesh out an interesting result that follows from the same information-theoretic ideas as Theorems 3 and 12 and that might be of independent interest: A new quantum version of the de Finetti Theorem [28] that applies to general quantum states, and not only to permutation-symmetric states as the previous known versions [47, 72, 62, 78, 24, 51, 26, 65, 80, 21, 20]. Theorem 13 follows directly from this result, which we discuss in more detail in section 2.3.

### 2.2.3 Polynomial-time Approximation Scheme for graphs with low threshold rank by Lasserre Hierarchy

Another class of Hamiltonians for which we can show a product-state assignment provides a good approximation to the energy, which can moreover be found in polynomial time, are 2-local Hamiltonians whose interaction graph has low *threshold rank*  $\text{rank}_\lambda$ , defined as the number of eigenvalues with value larger than  $\lambda$ . For a regular graph with  $n$  vertices, the degree  $D$  is related to the threshold rank as  $1/D \leq \lambda^2 + \text{rank}_\lambda/n$ . Therefore for  $\lambda = o(1)$  and  $\text{rank}_\lambda = o(n)$ , Theorem 3 shows that product states give a good approximation to the ground energy. However this does not say anything about the difficulty of finding the best product-state configuration.

In [14] the Lasserre hierarchy was shown to give good approximations of 2-CSPs on graphs with low threshold rank (examples of graphs with small threshold rank include small-set expanders and hypercontractive graphs [14]). The main result of this section is essentially an extension of their result to the quantum case.

**Theorem 14.** *Let  $G$  be a graph, and  $H = \mathbb{E}_{(i,j) \sim G} H_{i,j}$  a 2-local Hamiltonian on  $n$  qudits with each  $\|H_{i,j}\| \leq 1$ . Then it is possible to estimate the ground-state energy of  $H$  to within additive error  $\epsilon$  in time  $\exp(\text{poly}(d/\epsilon) \text{rank}_{\text{poly}(\epsilon/d)}(G)) \cdot \text{poly}(n)$ . If there exists a measurement  $\{A_x\}$  (i.e.  $A_x \geq 0$  and  $\sum_x A_x = I$ ) and a set of permutations  $\pi_{i,j}$  such that each  $H_{i,j} = -\sum_x A_x \otimes A_{\pi_{i,j}(x)}$  then the algorithm will run in time  $\exp(\log(d/\epsilon)) \cdot \text{poly}(n)$ .*

In fact the algorithm also gives a product state  $|\varphi\rangle = |\varphi_1\rangle \otimes \cdots \otimes |\varphi_n\rangle$  that achieves this error; i.e. such that

$$\langle \varphi | H | \varphi \rangle \leq e_0(H) + \epsilon. \quad (15)$$

The proof of Theorem 14 begins by using the Lasserre hierarchy for lower-bounding the ground-state energy of a Hamiltonian. Since variational methods provide upper bounds to the ground-state energy, such lower bounds can be useful even as heuristics, without any formal proof about their rate of convergence. Indeed, the SDP we discuss has been proposed as such a heuristic previously, by [15, 16] (building on previous work by [33, 56, 82, 54]). The rough idea is to approximate the minimum energy of an  $l$ -local Hamiltonian by minimizing over all locally compatible sets of  $k$ -body reduced density matrices (for some  $k \geq l$ ) satisfying a global positive

semidefinite condition. This optimization scheme is an SDP of size proportional to  $\binom{n}{k}d^k$ , and as  $k \rightarrow n$ , it approaches full diagonalization of the Hamiltonian <sup>8</sup>.

To produce the explicit product-state approximation, as well as the accuracy guarantee, we will need to use a rounding scheme due to [14], where the Lasserre hierarchy was similarly used for classical 2-CSPs. Indeed, the structure of the algorithm and proof of Theorem 14 follows [14] closely. In section 9 we give the details of applying the Lasserre hierarchy to quantum Hamiltonians and give the proof of Theorem 14.

## 2.3 de Finetti Theorems with no Symmetry

The de Finetti theorem says that the marginal probability distribution on  $l$  subsystems of a permutation-symmetric probability distribution on  $k \geq l$  subsystems is close to a convex combination of independent and identically distributed (i.i.d.) probability measures [28]. It allows us to infer a very particular form for the  $l$ -partite marginal distribution merely based on a symmetry assumption on the global distribution. Quantum versions of the de Finetti theorem state that a  $l$ -partite quantum state  $\rho_l$  that is a reduced state of a permutation-symmetric state on  $k \geq l$  subsystems is close (for  $k \gg l$ ) to a convex combination of i.i.d. quantum states, i.e.  $\rho_l \approx \int \mu(d\sigma) \sigma^{\otimes l}$  for a probability measure  $\mu$  on quantum states [51, 26, 65, 20, 31]. Quantum versions of the de Finetti theorem have found widespread applications in quantum information theory. However a restriction of all known de Finetti theorems (quantum and classical) is that they only apply to permutation-symmetric states. Is there any way of formulating a more general version of the theorem that would apply to a larger class of states? Here we give one such possible generalization.

We first show a new classical de Finetti theorem for general probability distributions.

**Theorem 15.** *Let  $p^{X_1, \dots, X_n}$  be a probability distribution on  $\Sigma^n$ , for a finite set  $\Sigma$ . Let  $\vec{i} = (i_1, \dots, i_k)$ ,  $\vec{j} = (j_1, \dots, j_m)$  be random disjoint ordered subsets of  $[n]$ , and let  $\vec{x} = (x_1, \dots, x_m)$  be distributed according to  $p^{X_{\vec{j}}} := p^{X_{j_1} \dots X_{j_m}}$ . Define  $p_{X_{\vec{j}}=\vec{x}}$  to be the conditional distribution resulting from taking  $X_{j_1} = x_1, \dots, X_{j_m} = x_m$ , and define  $p_{X_{\vec{j}}=\vec{x}}^{X_{\vec{i}}}$  to be the  $X_{i_1}, \dots, X_{i_k}$  marginal of this distribution.*

*Then for every integer  $t \leq n - k$  there is an integer  $m \leq t$  such that*

$$\mathbb{E}_{\substack{\vec{j}=(j_1, \dots, j_m) \\ \vec{x}=(x_1, \dots, x_m)}} \mathbb{E}_{\vec{i}=(i_1, \dots, i_k)} \left\| p_{X_{\vec{j}}=\vec{x}}^{X_{\vec{i}}} - p_{X_{\vec{j}}=\vec{x}}^{X_{i_1}} \otimes \dots \otimes p_{X_{\vec{j}}=\vec{x}}^{X_{i_k}} \right\|_1^2 \leq \frac{2k^2 \ln |\Sigma|}{t}, \quad (16)$$

In words the theorem says that given a probability distribution  $p^{X_1, \dots, X_n}$  over  $n$  variables, after conditioning on the values of at most  $t$  variables chosen uniformly at random, the remaining  $k$ -partite marginal distributions are close to a *product* distribution up to error  $O(k^2 \ln |\Sigma|/t)$ , on average over the choice of the  $k$ -partite marginal distribution, the variables  $(X_{j_1}, \dots, X_{j_m})$  that are conditioned on, and the outcomes  $(x_1, \dots, x_m)$  observed.

We can recover a version of the standard de Finetti theorem directly from Theorem 15. Let  $p^{X_1, \dots, X_n}$  be permutation symmetric. Then setting  $t = n - k$  and using convexity of the trace norm

<sup>8</sup>This is not to be confused with a more sophisticated SDP based on the noncommutative Positivstellensatz [30, 60, 57], which we will not analyze.

and the  $x^2$  function we find there is a measure  $\mu$  on distributions over  $\Sigma$  such that <sup>9</sup>

$$\left\| p^{X_1, \dots, X_k} - \int \mu(dq) q^{\otimes k} \right\|_1 \leq \sqrt{\frac{2k^2 \ln |\Sigma|}{n - k}}. \quad (17)$$

We now turn to the problem of obtaining a quantum version of Theorem 15. It turns out that it is straightforward to obtain one, unlike other quantum generalizations of the standard de Finetti theorem that require substantially more work (see [47, 72, 62, 78, 24, 51, 26, 65, 80, 21, 20, 31]). The only ingredient beyond Theorem 15 is the idea of applying a product measurement to a multipartite quantum state in order to turn it into a multipartite probability distribution (this will also be a central idea in the proof of the other theorems). If we want to translate distance bounds on the subsystems of the probability distribution into bounds on the original quantum state, then we will need an *informationally complete* measurement.

We model a quantum measurement as a quantum-classical channel  $\Lambda : \mathcal{D}(Q) \rightarrow \mathcal{D}(X)$ ,

$$\Lambda(\rho) = \sum_{x=1}^{|X|} \text{tr}(M_x \rho) |x\rangle \langle x|, \quad (18)$$

with  $\{M_x\}$  forming a POVM (positive-operator valued measure), i.e.  $0 \leq M_x$ ,  $\sum_x M_x = I$ , and  $\{|x\rangle\}$  forming an orthonormal basis. A measurement  $\Lambda$  is called *informationally complete* if the map  $\Lambda$  is injective; i.e. if any two density matrices can be distinguished by their classical measurement outcomes. We say that the *distortion* of a measurement  $\Lambda$  is  $\sup_{\xi \neq 0} \|\xi\|_1 / \|\Lambda(\xi)\|_1$ . There are several constructions of informationally complete measurements; for instance Ref. [51] gives a construction for a measurement in  $\mathcal{D}(\mathbb{C}^d)$  with distortion  $\sqrt{2}d^3$ .

If  $\Lambda$  is informationally complete, then it is straightforward to show that  $\Lambda^{\otimes k}$  is as well. However, we will need explicit bounds on the associated distortion.

**Lemma 16** (Informationally complete measurements). *For every positive integer  $d$  there exists a measurement  $\Lambda$  with  $\leq d^8$  outcomes such that for every positive integer  $k$  and every traceless operator  $\xi$ , we have*

$$\frac{1}{(18d)^{k/2}} \|\xi\|_1 \leq \|\Lambda(\xi)\|_1 \quad (19)$$

The proof is in Section 3.1.

**Theorem 17.** *Let  $\rho^{Q_1, \dots, Q_n} \in \mathcal{D}(Q^{\otimes n})$ . For every integer  $t \leq n - k$ , there is an integer  $m \leq t$  such that the following holds. Let  $\Lambda$  be the measurement from Lemma 16. Let  $\vec{i} = (i_1, \dots, i_k)$ ,  $\vec{j} = (j_1, \dots, j_m)$  be random disjoint ordered subsets of  $[n]$  and  $\vec{x} = (x_1, \dots, x_m)$  be the measurement outcomes resulting from applying  $\Lambda^{\otimes m}$  to  $Q_{\vec{j}} := Q_{j_1} \otimes \dots \otimes Q_{j_m}$ . Let  $\rho_{\vec{j}, \vec{x}}$  be the post-measurement state, conditioned on obtaining outcomes  $\vec{x}$ . Then*

$$\mathbb{E}_{\substack{\vec{j}=(j_1, \dots, j_m) \\ \vec{x}=(x_1, \dots, x_m)}} \mathbb{E}_{\vec{i}=(i_1, \dots, i_k)} \left\| \rho_{\vec{j}, \vec{x}}^{Q_{i_1} \dots Q_{i_k}} - \rho_{\vec{j}, \vec{x}}^{Q_1} \otimes \dots \otimes \rho_{\vec{j}, \vec{x}}^{Q_k} \right\|_1^2 \leq \frac{4 \ln(d) (18d)^k k^2}{t}, \quad (20)$$

---

<sup>9</sup>Note however we do know de Finetti theorems with a better error term. For instance, Diaconis and Freedman proved that the error in the R.H.S. of Eq. (17) can be improved to  $O(k^2/n)$  [28].



We note that a variant of Theorem 17 had been conjectured by Gharibian, Kempe and Regev also in the context of giving a polynomial-time approximation scheme for the local Hamiltonian problem on dense instances [38].

It is instructive to compare Theorem 17 with the usual previous method for applying quantum de Finetti theorems to quantum states without any symmetry. For example, in quantum key distribution (QKD), the sender and receiver treat each of the  $n$  transmitted qubits identically, but nothing forces the eavesdropper to use a symmetric attack. However, by randomly selecting a subset of  $k$  qubits, or equivalently, randomly permuting the  $n$  qubits and selecting the first  $k$ , we can obtain a state which is a reduction of a  $n$ -partite permutation-symmetric state. This strategy, together with the de Finetti theorem of Ref. [26], yields the bound

$$\left\| \mathbb{E}_{i_1, \dots, i_k} \rho^{Q_{i_1} \dots Q_{i_k}} - \int \mu(d\sigma) \sigma^{\otimes k} \right\|_1 \leq \frac{2d^2 k}{n - k} \quad (21)$$

Let  $\vec{j} = [n] \setminus \vec{i}$ , denote the  $n - k$  systems that are discarded. Then we can replace the partial trace on  $Q_{j_1}, \dots, Q_{j_m}$  with measurement (as indeed is performed by the proof of [26]) to rewrite Eq. (21) as

$$\left\| \mathbb{E}_{\vec{i}, \vec{j}, \vec{x}} \rho_{\vec{j}, \vec{x}}^{Q_{i_1} \dots Q_{i_k}} - \int \mu(d\sigma) \sigma^{\otimes k} \right\|_1 \leq \frac{2d^2 k}{n - k} \quad (22)$$

We can thus see that the advantage of Eq. (20) is that it allows the expectation over  $\vec{i}, \vec{j}, \vec{x}$  to be moved outside of the norm. This is what allows us to control the error when dealing with local Hamiltonians in which the interaction terms are not all identical. However, the cost of doing so blows up the error by a factor of  $d^k$ . It is an interesting question whether Theorem 17 can be improved to give a bound that only depends on  $\text{poly}(d, k)$ , and not on  $d^k$ .

For an approximation guarantee based on trace norm, there are examples showing that the error cannot be lower than  $\text{poly}(d, k, 1/n)$ . However, we can also measure the distance to product using other norms. One example is the LO (local-operations) norm, which is defined on a  $k$ -partite system as the maximum distinguishability achievable by local measurements and one-way communication to a referee; i.e.

$$\frac{1}{2} \|\xi\|_{\text{LO}} := \max \left\{ \sum_{x_1, \dots, x_k} \lambda_{x_1, \dots, x_k} \text{tr}(\xi(M_{x_1}^1 \otimes \dots \otimes M_{x_k}^k)) : 0 \leq \lambda_{x_1, \dots, x_k} \leq 1, \sum_x M_x^j = I. \right\}$$

Can we replace the 1-norm on the LHS of Eq. (20) with the LO norm and remove the  $(18d)^k$  term on the RHS of Eq. (20)? Such a result would be a simultaneous generalization of Theorem 17 above and Theorem 2 of our companion paper [20].

### 3 Quantum information theory background

In this section we review some necessary tools from quantum information theory. State tomography (i.e. the use of an informationally complete measurement) is discussed in Section 3.1. The use of quantum and classical entropies to quantify correlation is discussed in Section 3.2.

### 3.1 Informationally complete measurement

We prove Lemma 16, which states that there exists a measurement  $\Lambda$  on  $d$  dimensions with  $\leq d^8$  outcomes such that  $\Lambda^{\otimes k}$  has distortion  $\leq (18d)^{k/2}$ .

*Proof.* Suppose  $\{p_x, |\varphi_x\rangle\}$  is a 4-design, meaning that  $\{p_x\}$  is a probability distribution, each  $|\varphi_x\rangle$  is a pure state and  $\sum_x p_x \varphi_x^{\otimes 4}$  is the maximally mixed state on the symmetric subspace of  $(\mathbb{C}^d)^{\otimes 4}$  (see [41] for background on the symmetric subspace and on  $t$ -designs). Eq. (4) of [53] (see also Thm 9 of [55]) proved that if  $\Lambda(\xi) := \sum_x d p_x \langle \varphi_x | \xi | \varphi_x \rangle |x\rangle\langle x|$  then Eq. (19) holds. It remains only to show that 4-designs exist with size  $\leq d^8$ . But this follows from Carathéodory's theorem, since the space of traceless Hermitian matrices on the symmetric subspace has dimension  $\binom{d+3}{4}^2 - 1 \leq d^8 - 1$ .  $\square$

### 3.2 Multipartite mutual information

For a state  $\rho^{Q_1 \dots Q_k}$  we define the multipartite mutual information as

$$I(Q_1 : \dots : Q_k) := S(\rho^{Q_1 \dots Q_k} \| \rho^{Q_1} \otimes \dots \otimes \rho^{Q_k}) = S(Q_1) + \dots + S(Q_k) - S(Q_1 \dots Q_k). \quad (23)$$

For a quantum-classical state  $\rho^{Q_1 \dots Q_k R} = \sum_i p_i \rho_i^{Q_1 \dots Q_k} \otimes |i\rangle\langle i|_R$  we define the conditional multipartite mutual information as follows

$$I(Q_1 : \dots : Q_k | R)_\rho := \sum_i p_i I(Q_1 : \dots : Q_k)_{\rho_i}. \quad (24)$$

The multipartite mutual information satisfies the following properties:

**Lemma 18.**

1. *Chain Rule:*

$$I(A : BR) = I(A : R) + I(A : B | R). \quad (25)$$

2. *Multipartite-to-Bipartite [81]:*

$$I(Q_1 : \dots : Q_k | R) = I(Q_1 : Q_2 | R) + I(Q_1 Q_2 : Q_3 | R) + \dots + I(Q_1 \dots Q_{k-1} : Q_k | R). \quad (26)$$

3. *Monotonicity under Local Operations:* If  $\pi^{Q_1 \dots Q_k R} = \Lambda^{Q_1} \otimes \text{id}^{Q_2 \dots Q_k}(\rho^{Q_1 \dots Q_k R})$ , then

$$I(Q_1 : \dots : Q_k | R)_\pi \leq I(Q_1 : \dots : Q_k | R)_\rho \quad (27)$$

4. *QC monotonicity:* For any measurement  $\Lambda^{Q \rightarrow X}$  and any state  $\rho^{Q_1 Q_2}$ , we have:

$$I(X_1 : X_2)_{(\Lambda^{Q_1 \rightarrow X_1} \otimes \Lambda^{Q_2 \rightarrow X_2})(\rho)} \leq I(Q_1 : Q_2)_{(\text{id} \otimes \Lambda^{Q_2 \rightarrow X_2})(\rho)} \leq S(Q_1)_\rho. \quad (28)$$

5. *Pinsker's Inequality:*

$$I(Q_1 : \dots : Q_k)_\rho \geq \frac{1}{2} \|\rho^{Q_1 \dots Q_k} - \rho^{Q_1} \otimes \dots \otimes \rho^{Q_k}\|_1^2. \quad (29)$$

6. *Upper Limit:* For any state  $\rho^{Q_1 Q_2}$  and measurements  $\Lambda_1, \Lambda_2$ ,

$$I(Q_1 : Q_2)_\rho \leq 2 \min(\ln |Q_1|, \ln |Q_2|). \quad (30)$$

$$I(X_1 : X_2)_{(\Lambda_1^{Q_1 \rightarrow X_1} \otimes \Lambda_2^{Q_2 \rightarrow X_2})(\rho)} \leq \min(\ln |Q_1|, \ln |Q_2|). \quad (31)$$

## 4 Self-Decoupling Lemmas

The main technical tool in this paper is an information-theoretic bound that we call “self-decoupling” and which is based on a very similar technique in [63]. Decoupling refers to the situation when conditioning on one random variable leaves some others nearly decoupled (i.e. independent). This can be quantified by saying that the conditional mutual information is low. We use the term “self-decoupling” to mean that a collection of random variables can be approximately decoupled by conditioning on a small number of those variables. We learned about this technique from [63], which introduced it in order to analyze SDP hierarchies. Our companion paper [20] applied the same ideas to prove new quantum de Finetti theorems.

For a distribution  $\mu$  on  $[n]$ , define the distribution  $\mu^{\wedge m}$  to be the distribution on  $[n]^m$  obtained by sampling  $m$  times without replacement according to  $\mu$ ; i.e.

$$\mu^{\wedge m}(i_1, \dots, i_m) = \begin{cases} 0 & \text{if } i_1, \dots, i_m \text{ are not all distinct} \\ \frac{\mu(i_1) \cdots \mu(i_m)}{\sum_{j_1, \dots, j_m \text{ distinct}} \mu(j_1) \cdots \mu(j_m)} & \text{otherwise} \end{cases} \quad (32)$$

Note that  $\mu^{\wedge m}$  is only a valid probability distribution if  $m \leq |\text{supp } \mu|$ .

**Lemma 19** (Self-Decoupling Lemma). *Let  $X_1, \dots, X_n$  be classical random variables with some arbitrary joint distribution, let  $\mu$  be a distribution on  $[n]$  and  $k < |\text{supp } \mu|$ . Then*

$$\mathbb{E}_{0 \leq k' < k} \mathbb{E}_{(a, b, c_1, \dots, c_{k'}) \sim \mu^{\wedge k'+2}} I(X_a : X_b | X_{c_1} \dots X_{c_{k'}}) \leq \frac{1}{k} \mathbb{E}_{i \sim \mu} I(X_i : X_i^c), \quad (33)$$

where  $X_i^c := X_{[n] \setminus \{i\}}$ .

*Proof.* Sample  $(i, j_1, \dots, j_k) \sim \mu^{\wedge k+1}$ . Then

$$\frac{1}{k} \mathbb{E}_{i \sim \mu} I(X_i : X_i^c) \geq \frac{1}{k} \mathbb{E}_{(i, j_1, \dots, j_k) \sim \mu^{\wedge k+1}} I(X_i : X_{j_1} \dots X_{j_k}) \quad \text{monotonicity} \quad (34)$$

$$= \mathbb{E}_{(i, j_1, \dots, j_k) \sim \mu^{\wedge k+1}} \mathbb{E}_{0 \leq k' < k} I(X_i : X_{j_{k'+1}} | X_{j_1} \dots X_{j_{k'}}) \quad \text{chain rule} \quad (35)$$

$$= \mathbb{E}_{0 \leq k' < k} \mathbb{E}_{(a, b, c_1, \dots, c_{k'}) \sim \mu^{\wedge k'+2}} I(X_a : X_b | X_{c_1} \dots X_{c_{k'}}) \quad (36)$$

In the last line we have relabeled  $i \rightarrow a$ ,  $j_{k'+1} \rightarrow b$  and  $j_1, \dots, j_{k'} \rightarrow c_1, \dots, c_{k'}$ .  $\square$

We can also derandomize Lemma 19 to obtain:

**Corollary 20** (Derandomized Self-Decoupling). *Let  $X_1, \dots, X_n$  be classical random variables with joint distribution  $p$ , let  $\mu$  be a distribution on  $[n]$  and  $k < |\text{supp } \mu|$ . Then there exists  $k' < k$  and  $c_1, \dots, c_{k'} \in [n]$  all distinct such that*

$$\mathbb{E}_{\substack{(a, b) \sim \mu^{\wedge 2} \\ a, b \notin \{c_1, \dots, c_{k'}\}}} I(X_a : X_b | X_{c_1}, \dots, X_{c_{k'}}) \leq \frac{1}{k} \mathbb{E}_{i \sim \mu} I(X_i : X_i^c). \quad (37)$$

The proof is immediate.

In Lemma 19, we needed  $b, c_1, \dots, c_k$  to be drawn from an exchangeable distribution, but did not make use of the fact that  $a$  was also drawn from the same distribution. Thus, it is possible to generalize the argument to the case where  $a$  and  $b, c_1, \dots, c_k$  index entirely different sets of random variables. A further generalization is to add a random variable  $Z$  upon which we condition all of the other variables.

**Lemma 21** (Bipartite Self-Decoupling Lemma). *Let  $X_1, \dots, X_m, Y_1, \dots, Y_n, Z$  be jointly distributed random variables. Let  $\sigma$  be a probability distribution over another random variable  $W$ . For each value  $w$  in the support of  $W$ , define distributions  $\mu_w$  on  $[m]$  and  $\nu_w$  on  $[n]$ . Choose  $k < \min_w |\text{supp } \nu_w|$ .*

$$\mathbb{E}_{0 \leq k' < k} \mathbb{E}_{w \sim \sigma} \mathbb{E}_{a \sim \mu_w} \mathbb{E}_{(b, c_1, \dots, c_{k'}) \sim \nu_w^{\wedge k' + 1}} I(X_a : Y_b | Y_{c_1} \dots Y_{c_{k'}}, Z) \leq \frac{1}{k} \mathbb{E}_{w \sim \sigma} \mathbb{E}_{i \sim \mu_w} I(X_i : \cup_{j \in \text{supp } \nu_w} Y_j | Z) \quad (38)$$

*Proof.*

$$\frac{1}{k} \mathbb{E}_{w \sim \sigma} \mathbb{E}_{i \sim \mu_w} I(X_i : \cup_{j \in \text{supp } \nu_w} Y_j | Z) \geq \frac{1}{k} \mathbb{E}_{w \sim \sigma} \mathbb{E}_{i \sim \mu_w} \mathbb{E}_{j_1, \dots, j_k \sim \nu_w^{\wedge k}} I(X_i : Y_{j_1} \dots Y_{j_k} | Z) \quad (39)$$

$$= \mathbb{E}_{w \sim \sigma} \mathbb{E}_{i \sim \mu_w} \mathbb{E}_{j_1, \dots, j_k \sim \nu_w^{\wedge k}} \mathbb{E}_{0 \leq k' < k} I(X_i : Y_{j_{k'+1}} | Y_{j_1} \dots Y_{j_k}, Z) \quad (40)$$

$$= \mathbb{E}_{0 \leq k' < k} \mathbb{E}_{w \sim \sigma} \mathbb{E}_{a \sim \mu_w} \mathbb{E}_{(b, c_1, \dots, c_{k'}) \sim \nu_w^{\wedge k' + 1}} I(X_a : Y_b | Y_{c_1} \dots Y_{c_{k'}}, Z) \quad (41)$$

As in the proof of Lemma 19, (39), (40) and (41) follow in turn from monotonicity, the chain rule and relabeling variables.  $\square$

The main example where we will use Lemma 21 is to analyze the case when  $Y_1, \dots, Y_n$  are variables that have been divided into blocks. Suppose we have partitioned  $[n]$  into  $n/m$  blocks  $V_1, \dots, V_{n/m}$  with  $V_i = \{v_{i,1}, \dots, v_{i,m}\}$ . Then define the “coarse-grained” variables  $X_i = (Y_{v_{i,1}}, \dots, Y_{v_{i,m}})$ . Suppose further that we have a distribution  $\nu$  over  $[n]$  with an induced distribution over blocks  $\mu$ ; i.e.  $\mu(i) = \sum_{a \in V_i} \nu(a)$ . Then we can take  $\sigma$  to be a distribution over blocks  $w$  and define  $\mu_w$  to be  $\mu$  conditioned on not choosing  $w$  and  $\nu_w$  to be  $\nu$  conditioned on restricting to  $j \in V_w$ ; i.e.

$$\mu_w(i) = \begin{cases} 0 & \text{if } i = w \\ \frac{\mu(i)}{\sum_{i' \neq w} \mu(i')} & \text{otherwise} \end{cases} \quad \nu_w(a) = \begin{cases} 0 & \text{if } a \notin V_w \\ \frac{\nu(a)}{\sum_{a' \in V_w} \nu(a')} & \text{otherwise} \end{cases} \quad (42)$$

This will be useful in our proof of Theorem 6 where we are interested in the correlations between an individual system and an adjacent block of systems. For this application we will present a derandomized version of Lemma 21 specialized to this block scenario.

**Corollary 22** (Derandomized Block Self-Decoupling). *Let  $n, m, V_1, \dots, V_{n/m}, X_1, \dots, X_{n/m}, Y_1, \dots, Y_n, Z$  be defined as in the above paragraph. Let  $B(a)$  denote the identity of the block containing  $a$ . Then for any  $k < m$  there exists  $k' < k$  and sites  $C := \{c_j^i\}_{i \in [n/m], j \in [k']}$  such that  $c_j^i \in V_i$  and*

$$\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \notin (V_i \cup C)} I(X_i : Y_a | Y_{c_1^{B(a)}} \dots Y_{c_{k'}^{B(a)}}, Z) \leq \frac{1}{k} \mathbb{E}_{i \neq j} I(X_i : X_j | Z) \quad (43)$$

*Proof.* Invoking Lemma 21 with  $\sigma, \mu_w, \nu_w$  chosen as in Eq. (42) yields

$$\mathbb{E}_{0 \leq k' < k} \mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \notin (V_i \cup C)} \mathbb{E}_{\substack{c_1, \dots, c_{k'} \\ B(a) = B(c_1) = \dots = B(c_{k'})}} I(X_i : Y_a | Y_{c_1}, \dots, Y_{c_{k'}}, Z) \leq \frac{1}{k} \mathbb{E}_{i \neq j} I(X_i : X_j | Z) \quad (44)$$

For some choice of  $k'$  (which we fix from now on), we have

$$\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \notin (V_i \cup C)} \mathbb{E}_{\substack{c_1, \dots, c_{k'} \\ B(a) = B(c_1) = \dots = B(c_{k'})}} I(X_i : Y_a | Y_{c_1}, \dots, Y_{c_{k'}}, Z) \leq \frac{1}{k} \mathbb{E}_{i \neq j} I(X_i : X_j | Z) \quad (45)$$

Now, for each block  $j \neq i$ , choose  $c_1, \dots, c_{k'} \in V_i$  to minimize  $\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \in V_j} I(X_i : Y_a | Y_{c_1}, \dots, Y_{c_{k'}}, Z)$ . Call this choice  $c_1^j, \dots, c_{k'}^j$ . Again using the fact that the minimum of random variable is never greater than its expectation, we obtain Eq. (43).  $\square$

*Remark:* Can we generalize these results to quantum self-decoupling theorems? The chain rule of conditional mutual information works just as well for von Neumann entropies. So essentially the same argument (modified only by the fact that  $I(A : B)$  can be as large as  $2 \ln |A|$ ) works there, and was used already in [21]. However, what fails is the conditioning. Unlike for classical random variables, the quantum conditional mutual information cannot be written as an average of mutual informations of some conditional distribution. This raises a beautiful open problem, which is to understand the structure of quantum states that have very small conditional mutual information [44, 48, 21].

The strategy of our paper will instead be to apply the self-decoupling lemmas to the classical random variables resulting from measuring quantum states. Our companion paper [20] followed a similar strategy, sometimes leaving a system quantum if there was no chance of conditioning on it. In the language of this paper, this can be thought of as observing that in bipartite self-decoupling (Lemma 21), we could have taken  $X_1, \dots, X_m$  to be quantum systems.

## 5 Proof of main results: Theorems 3, 6 and 8

This section uses the background results from Section 3 and the decoupling lemmas from Section 4 to prove Theorems 6 (in Section 5.1) and 8 (in Section 5.2). To avoid redundancy, we give only a sketch of the proof of Theorem 3 in Section 5.3.

### 5.1 Regular partitioned graphs

In this section we complete the proof of Theorem 6.

**Theorem 6 (restatement).** *Let  $G = (V, E)$  be a  $D$ -regular graph, with  $n = |V|$ ,  $m$  a positive integer dividing  $n$  and  $\{V_1, \dots, V_{n/m}\}$  a partition of  $V$  into sets each of size  $m$ . Define  $\bar{\Phi}_G := \mathbb{E}_{i \in [n/m]} \Phi_G(V_i)$  to be the average expansion of the blocks.*

*Let  $\rho$  be a state on  $(\mathbb{C}^d)^{\otimes n}$  and define  $\bar{I} := \mathbb{E}_{i \in [n/m]} I(V_i : V_i^c)_\rho$  to be the average mutual information of  $V_i$  with its complement for state  $\rho$ .*

*Then there exists a state  $\sigma$  such that*

1.  $\sigma$  is separable with respect to the partition  $(V_1, \dots, V_{n/m})$ ; i.e.  $\sigma = \sum_z p_z \sigma_z$  for some states  $\sigma_z$  of the form  $\sigma_z = \sigma_z^{V_1} \otimes \dots \otimes \sigma_z^{V_{n/m}}$ ; and
- 2.

$$\mathbb{E}_{(i,j) \in E} \|\rho^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 \leq 9 \left( \frac{d^2 \bar{\Phi}_G}{D} \frac{\bar{I}}{m} \right)^{1/6} + \frac{1}{m} + \frac{m}{n} \quad (46)$$

*Proof.* Our proof will proceed by upper-bounding a series of different correlations, on the way to our final statement. The proof proceeds by bounding the following quantities in turn:

1. Classical block-block correlations.
2. Classical block-site correlations.
3. Classical block-site correlations along edges between blocks. (All later correlations will also be measured in terms of their average across edges between blocks.)
4. Classical site-site correlations.
5. Classical site-site correlations using variational distance.
6. Quantum site-site correlations using trace distance.
7. Constructing the separable approximating state.

1. *Classical block-block correlations.* Let  $p^{Y_1, \dots, Y_n} := \Lambda^{\otimes n}(\rho)$  be a probability distribution obtained by measuring all  $n$  subsystems of  $\rho$  with the informationally complete measurement  $\Lambda$  given by Lemma 16. Let  $X_1, \dots, X_{n/m}$  be the random variables associated with the partition  $V_1 \cup \dots \cup V_{n/m}$  of the sites; i.e. if  $V_i = \{v_{i,1}, \dots, v_{i,m}\}$ , then  $X_i := (Y_{v_{i,1}}, \dots, Y_{v_{i,m}})$ . By the monotonicity of mutual information under local operations (specifically under  $\Lambda$ ), we have  $I(X_i : X_i^c)_p \leq I(V_i : V_i^c)_\rho$ . (We remark that in Corollary 7 we have  $\rho = \psi_0$  and have replaced  $I(V_i : V_i^c)_\rho$  with  $S(V_i)_{\psi_0}$ . This is a consequence of part 4 of Lemma 18.)

Let  $\mu$  be the uniform distribution on  $[n/m]$  and let  $k_1 = \frac{n}{m} \delta_1$  where  $\delta_1$  is a parameter we will set later. By Corollary 20, there exists  $k' < k_1$  and  $c_1, \dots, c_{k'}$  such that

$$\mathbb{E}_{\substack{(i,j) \sim \mu^{\wedge 2} \\ i,j \notin \{c_1, \dots, c_{k'}\}}} I(X_i : X_j | X_{c_1}, \dots, X_{c_{k'}}) \leq \frac{m}{n \delta_1} \mathbb{E}_i I(V_i : V_i^c)_\rho = \frac{m \bar{I}}{n \delta_1 \delta_2}. \quad (47)$$

2. *Classical block-site correlations.* Define  $Z_1 := X_{c_1} \dots X_{c_{k'}}$  and let  $B(a)$  denote the identity of the block containing  $a$ . Choose  $k_2 = m \delta_2$  for some  $\delta_2$  we will set later. Now we invoke Corollary 22 to show that there exists  $k'' < k_2$  and  $\{c_j^i\}_{i \in [n/m], j \in [k'']}$  such that

$$\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \notin V_i} I(X_i : Y_a | Y_{c_1^{B(a)}}, \dots, Y_{c_{k'}^{B(a)}}) \leq \frac{1}{m \delta_2} \mathbb{E}_{i \neq j} I(X_i : X_j | Z_1) \quad (48a)$$

$$\leq \frac{\bar{I}}{n \delta_1 \delta_2} \quad (48b)$$

Define  $Z_2 := Y_{c_1^1} \dots Y_{c_{k''}^{n/m}}$  and define  $Z := Z_1 Z_2$ . Then (48) becomes

$$\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \notin V_i} I(X_i : Y_a | Z) \leq \frac{\bar{I}}{n \delta_1 \delta_2}. \quad (49)$$

3. *Classical block-site correlations along edges between blocks.* Define  $\Gamma(i)$  to be the multiset of sites neighboring  $V_i$ ; i.e.  $\Gamma(i)$  is the image of the map  $(a, b) \mapsto a$  applied to each  $(a, b) \in E$  with  $b \in V_i$ . For each  $i$ ,  $|\Gamma(i)| = mD\Phi_G(V_i)$ , and

$$\sum_{i \in [n/m]} |\Gamma(i)| = mD \sum_{i \in [n/m]} \Phi_G(V_i) = nD\bar{\Phi}_G.$$

On the other hand, the expectation on the LHS of Eq. (49) is over  $\leq (n/m) \cdot n$  pairs of  $i, a$ , so we can use it to upper-bound the expectation over  $i \in [n/m], a \in \Gamma(i)$

$$\mathbb{E}_{i \in [n/m]} \mathbb{E}_{a \in \Gamma(i)} I(X_i : Y_a | Z) \leq \frac{\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}. \quad (50)$$

4. *Classical site-site correlations.* For a pair of sites  $a, b$  we will define the relation  $a \sim b$  similarly; i.e. let  $a \sim b$  if  $(a, b) \in E$  and  $B(a) \neq B(b)$ . By monotonicity of mutual information under partial trace, we have that  $I(Y_a : Y_b | Z) \leq I(X_{B(a)} : Y_b | Z)$ . Also note that the uniform distribution over  $i$  is also obtained by choosing  $a$  random and setting  $i = B(a)$ . Then

$$\mathbb{E}_{a \sim b} I(Y_a : Y_b | Z) \leq \frac{\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}. \quad (51)$$

5. *Classical site-site correlations using variational distance.* Since the (classical) conditional mutual information is an average of mutual informations of conditional distributions, Eq. (51) can be rewritten as

$$\mathbb{E}_{z \sim p^Z} \mathbb{E}_{a \sim b} I(Y_a : Y_b)_{q_z} \leq \frac{\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}. \quad (52)$$

where we have defined  $q_z := p|_{Z=z}$ . Now we can use Pinsker's inequality and convexity of  $x \mapsto x^2$  to obtain

$$\mathbb{E}_{z \sim p^Z} \mathbb{E}_{a \sim b} \|q_z^{Y_a Y_b} - q_z^{Y_a} \otimes q_z^{Y_b}\|_1^2 \leq \frac{2\bar{I}}{m \delta_1 \delta_2 D \bar{\Phi}_G}. \quad (53)$$

$$\mathbb{E}_{z \sim p^Z} \mathbb{E}_{a \sim b} \|q_z^{Y_a Y_b} - q_z^{Y_a} \otimes q_z^{Y_b}\|_1 \leq \sqrt{\frac{2\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}}. \quad (54)$$

6. *Quantum site-site correlations using trace distance.* Define  $C := \bigcup_{i \in [k']} V_i \cup \{c_j^i : i \in [n/m], j \in [k'']\}$ . Let  $\tau_z^{Q_1 \dots Q_n}$  be the state that results from measuring (using  $\Lambda$ ) only the sites in  $C$  and obtaining outcome  $z$ . Observe that  $z$  has distribution  $p^C$  and that

$$\text{tr}_C \rho = \mathbb{E}_z \text{tr}_C \tau_z = \sum_z p^C(z) \text{tr}_C \tau_z. \quad (55)$$

Then Eq. (54) and Lemma 16 imply that

$$\mathbb{E}_{z \sim p^Z} \mathbb{E}_{a \sim b} \|\tau_z^{Q_a Q_b} - \tau_z^{Q_a} \otimes \tau_z^{Q_b}\|_1 \leq 18d \sqrt{\frac{2\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}}. \quad (56)$$

This can already be seen as a useful intermediate result: measuring a small fraction of the systems leaves a state that is close to product across an average edge.

7. *Constructing the separable approximating state.* Now define  $\sigma_z := \tau_z^{V_1} \otimes \dots \otimes \tau_z^{V_{n/m}}$ ; i.e.  $\sigma_z$  is the same as  $\tau_z$  within blocks, but has no entanglement between blocks. Let  $\sigma_z = \sum_z p^G(z) \sigma$ . Our final goal is to bound

$$\mathbb{E}_{(a,b) \in E} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1. \quad (57)$$

We do so by considering separately three types of edges: those that are across blocks, those that are within blocks but contain a vertex in  $C$ , and the rest. Their definitions, and cardinalities are as follows:

$$\begin{aligned} E_{\text{across}} &:= \{(a, b) : a \sim b\} = \{(a, b) \in E : B(a) \neq B(b)\} \\ E_{\text{measured}} &:= \{(a, b) \in E : B(a) = B(b), \{a, b\} \cap C \neq \emptyset\} \\ E_{\text{within}} &:= E - E_{\text{across}} - E_{\text{measured}} \end{aligned}$$

We can bound their cardinalities (using  $|E| = nD/2$ ) by

$$|E_{\text{across}}| = \frac{1}{2} nD \bar{\Phi} = \bar{\Phi} |E| \quad (58a)$$

$$|E_{\text{measured}}| \leq |C|D \leq nD(\delta_1 + \delta_2) = 2(\delta_1 + \delta_2)|E|. \quad (58b)$$

For  $(a, b) \in E_{\text{within}}$ , we have  $\sigma^{Q_a Q_b} = \rho^{Q_a Q_b}$ . Thus the only contributions to Eq. (57) can come from  $E_{\text{across}}$  and  $E_{\text{measured}}$ . The latter average we can bound simply by

$$\mathbb{E}_{(a,b) \in E_{\text{measured}}} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 \leq 2. \quad (59)$$

For edges across blocks, we recall that  $\sigma_z^{Q_a Q_b} = \tau_z^{Q_a} \otimes \tau_z^{Q_b}$ . And for edges that do not intersect  $C$ , we have  $\rho^{Q_a Q_b} = \mathbb{E}_z \tau_z^{Q_a Q_b}$  (see Eq. (55)). Thus, we obtain

$$\begin{aligned} \mathbb{E}_{(a,b) \in E_{\text{across}}} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 &= \mathbb{E}_{(a,b) \in E_{\text{across}}} \left\| \mathbb{E}_z \tau_z^{Q_a Q_b} - \mathbb{E}_z \tau_z^{Q_a} \otimes \tau_z^{Q_b} \right\|_1 \\ &\leq \mathbb{E}_{(a,b) \in E_{\text{across}}} \mathbb{E}_z \|\tau_z^{Q_a Q_b} - \tau_z^{Q_a} \otimes \tau_z^{Q_b}\|_1 \quad \text{triangle inequality} \\ &\leq 18d \sqrt{\frac{2\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}}. \quad \text{using Eq. (56)} \end{aligned}$$

Putting everything together we have

$$\mathbb{E}_{(a,b) \in E} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 \leq \bar{\Phi}_G 18d \sqrt{\frac{2\bar{I}/m}{\delta_1 \delta_2 D \bar{\Phi}_G}} + (\delta_1 + \delta_2). \quad (60)$$

We balance these terms by setting  $\delta_1, \delta_2$  to be the smallest numbers greater than  $(18d)^{1/3} (2\bar{\Phi}_G \bar{I}/mD)^{1/6}$  for which  $\frac{n}{m}\delta_1$  and  $m\delta_2$  are integers. Thus we obtain

$$\mathbb{E}_{(a,b) \in E} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 \leq 3(18\sqrt{2})^{1/3} \left( \frac{d^2 \bar{\Phi}_G \bar{I}/m}{D} \right)^{1/6} + \frac{1}{m} + \frac{m}{n}. \quad (61)$$

The numerical constant  $3(18\sqrt{2})^{1/3} = 8.825 \dots \leq 9$ .  $\square$



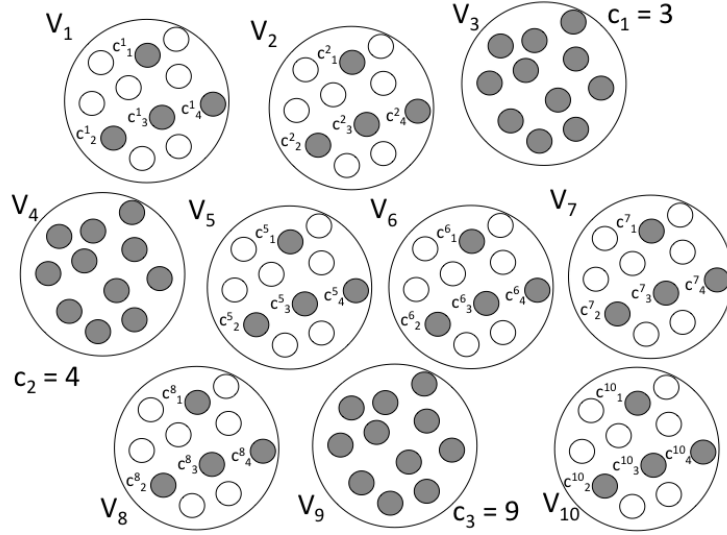


Figure 1: Construction used in the proof of Theorem 6: In the first step  $k'$  blocks  $c_1, \dots, c_{k'}$  are chosen at random and we condition on the values obtained by measuring  $\Lambda$  on all the sites belonging to  $V_{c_1} \cup \dots \cup V_{c_{k'}}$ . In the second step we condition on the values obtained by measuring  $\Lambda$  on sites  $\{c_j^i : i \in [n/m], j \in [k']\}$ , where  $c_j^i \in V_i$ . In the picture the conditioned sites are denoted by filled disks and the unconditioned sites by unfilled disks.

## 5.2 Variable degree

**Theorem 8 (restatement).** *Let  $G$  be a symmetric matrix with nonnegative entries summing to one and let  $A_{i,j} := G_{i,j} / \sum_{i'} G_{i',j}$ . Then for any  $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$  there exists a globally separable state  $\sigma$  such that*

$$\mathbb{E}_{(i,j) \sim G} \|\rho^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 \leq 14 (d^4 \ln(d) \operatorname{tr}[A^2] \|\pi\|_2^2)^{1/8} + \|\pi\|_2^2 \quad (62)$$

*If further we have a Hamiltonian  $H$  given by Eq. (10), then there exists a state  $|\varphi\rangle = |\varphi_1\rangle \otimes \dots \otimes |\varphi_n\rangle$  such that*

$$\operatorname{tr}(H\varphi) \leq e_0(H) + 14 (d^4 \ln(d) \operatorname{tr}[A^2] \|\pi\|_2^2)^{1/8} + \|\pi\|_2^2 \quad (63)$$

*Proof.* Let  $p^{X_1, \dots, X_n} := \Lambda^{\otimes n}(\rho)$  be a probability distribution obtained by measuring all  $n$  subsystems of  $\rho$  with the informationally complete measurement  $\Lambda$  given by Lemma 16. Observe that by Eq. (31),  $I(X_i : X_i^c) \leq \ln(d)$  for each  $i$ . Let  $\delta > 0$  be a constant we will choose later. Applying self-decoupling (Lemma 19) with  $\mu = \pi$  and  $k$  to be chosen later, we find

$$\mathbb{E}_{0 \leq k' < k} \mathbb{E}_{(a,b,c_1, \dots, c_{k'}) \sim \pi^{\wedge k'+2}} I(X_a : X_b | X_{c_1}, \dots, X_{c_{k'}}) \leq \frac{\ln(d)}{k}. \quad (64)$$

Define  $C := \{c_1, \dots, c_{k'}\}$  and define  $\gamma := \mathbb{E}_{a,b} I(X_a : X_b | X_C)_p$ , where the expectation is taken over  $a, b$  sampled from  $\pi^{\wedge 2}$  conditioned on  $a, b \notin C$ . Then Eq. (64) could be expressed succinctly as

$$\mathbb{E}_C[\gamma] \leq \ln(d)/k. \quad (65)$$

(Note that  $k'$  is determined by  $C$ , so  $\mathbb{E}_C$  represents the average over choices of both  $k'$  and  $c_1, \dots, c_{k'}$ .)

Abbreviate  $Z := X_{c_1} \dots X_{c_{k'}}$  and define

$$\Delta_z(a, b) := \begin{cases} \|p^{X_a X_b}|_{Z=z} - p^{X_a}|_{Z=z} \otimes p^{X_b}|_{Z=z}\|_1 & \text{if } a \neq b \text{ and } a, b \notin C \\ 0 & \text{otherwise} \end{cases} \quad (66)$$

By Pinsker's inequality,

$$\Delta_z(a, b) \leq \sqrt{2I(X_a : X_b|Z=z)_p}. \quad (67)$$

Define the inner product  $\langle A, B \rangle := \text{tr}(A^\dagger B)$ , and define  $\Delta_z^2$  to be the entrywise square of  $\Delta_z$ . Then (67) implies that

$$\mathbb{E}_{z \sim p^Z} \langle \pi \pi^T, \Delta_z^2 \rangle \leq 2\gamma. \quad (68)$$

We would like to show that  $\mathbb{E}_z \Delta_z(a, b)$  is small when we average it over  $(a, b) \sim G$ . However, Eq. (68) bounds the average over the distribution  $\pi \pi^T$ . To relate these, we will divide  $G$  into a component that is well approximated by  $\pi \pi^T$  and a component that isn't. For some parameter  $\lambda > 0$  to be determined later, define

$$B := \{(i, j) : G_{i,j} > \lambda \pi_i \pi_j\} = \{(i, j) : A_{i,j} > \lambda \pi_i\}.$$

We can then bound

$$\eta := \sum_{(i,j) \in B} G_{i,j} = \sum_{(i,j) \in B} A_{i,j} \pi_j \leq \sum_{(i,j) \in B} \frac{A_{i,j} A_{j,i}}{\lambda} \leq \frac{\text{tr}[A^2]}{\lambda}. \quad (69)$$

By the definition of  $\eta$ , we can find a probability distribution  $G'$  such that  $G \leq \lambda \pi \pi^T + \eta G'$ , where  $\leq$  is an element-wise inequality.

We will now bound the average correlation across edges. Define  $\Delta := \mathbb{E}_{z \sim p^Z} \Delta_z$ . Then

$$\langle G, \Delta \rangle \leq \sqrt{\langle G, \Delta^2 \rangle} \quad \text{Cauchy-Schwarz} \quad (70a)$$

$$\leq \sqrt{\langle \lambda \pi \pi^T, \Delta^2 \rangle + \langle \eta G', \Delta^2 \rangle} \quad (70b)$$

$$\leq \sqrt{2\gamma\lambda + 4\eta} \quad \text{using Eq. (68) and } \Delta(a, b) \leq 2 \quad (70c)$$

$$\leq \sqrt{2\gamma\lambda + 4 \frac{\text{tr}[A^2]}{\lambda}} \quad \text{using Eq. (69)} \quad (70d)$$

$$= (32\gamma \text{tr}[A^2])^{1/4} \quad \text{choosing } \lambda \text{ to balance the two terms.} \quad (70e)$$

Now we turn to approximating  $\rho$  with a mixture of product states. Define  $\tau$  to be the state resulting from measuring  $\rho$  only on systems  $Q_{c_1}, \dots, Q_{c_{k'}}$ . Let  $Q'$  denote the remaining  $n - k'$  quantum systems. Then we can write

$$\tau = \mathbb{E}_{z \sim p^Z} \tau_z^{Q'} \otimes |z\rangle\langle z|^Z,$$

and we have  $\rho^{Q'} = \tau^{Q'} = \mathbb{E}_{z \sim p^Z} \tau_z^{Q'}$ . It will be convenient to extend each  $\tau_z$  to be defined on all  $Q_1 \dots Q_n$  by defining

$$\tau_z^{Q_1 \dots Q_n} := \tau_z^{Q'} \otimes \bigotimes_{c \in C} (I/d)^{Q_c}.$$

(The state  $I/d$  here is an arbitrary choice; any other state would also work for the proof.) Now define

$$\sigma_z^{Q_1 \dots Q_n} := \tau_z^{Q_1} \otimes \dots \otimes \tau_z^{Q_n}, \quad (71)$$

and  $\sigma := \mathbb{E}_{z \sim p^Z} \sigma_z$ . Then

$$\begin{aligned} \mathbb{E}_{z \sim p^Z} \mathbb{E}_{(a,b) \sim G} \|\tau_z^{Q_a Q_b} - \sigma_z^{Q_a Q_b}\|_1 &= \mathbb{E}_{z \sim p^Z} \mathbb{E}_{(a,b) \sim G} \|\tau_z^{Q_a Q_b} - \tau_z^{Q_a} \otimes \tau_z^{Q_b}\|_1 \\ &\leq \mathbb{E}_{z \sim p^Z} \mathbb{E}_{(a,b) \sim G} 18d \Delta_z(a, b) && \text{using Lemma 16} \\ &\leq 18d (32\gamma \text{tr}[A^2])^{1/4} && \text{from Eq. (70)} \end{aligned}$$

Finally we evaluate the total error. One new challenge is that  $\pi(C)$  is not always  $\leq k/n$ . Observe that

$$\mathbb{E}_{(a,b) \sim G} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 \leq 2\pi(C) + \mathbb{E}_{(a,b) \sim G} \Delta(a, b) \quad (72)$$

$$\leq 2\pi(C) + 18d (32\gamma \text{tr}[A^2])^{1/4} \quad (73)$$

Now  $\mathbb{E}_C[\pi(C)] = \mathbb{E}_{k'}[k'] \|\pi\|_2^2 \leq \frac{1}{2}k \|\pi\|_2^2$ . Choose  $C$  to minimize Eq. (73). This yields

$$\begin{aligned} \mathbb{E}_{(a,b) \sim G} \|\rho^{Q_a Q_b} - \sigma^{Q_a Q_b}\|_1 &\leq \mathbb{E}_C \left[ 2\pi(C) + 18d (32\gamma \text{tr}[A^2])^{1/4} \right] && \text{since min} \leq \text{expectation} \\ &\leq k \|\pi\|_2^2 + 18d \left( 32 \mathbb{E}_C[\gamma] \text{tr}[A^2] \right)^{1/4} && \text{convexity of } x \mapsto x^4 \\ &\leq k \|\pi\|_2^2 + 18d \left( 32 \frac{\ln(d) \text{tr}[A^2]}{k} \right)^{1/4} && \text{from Eq. (65)} \\ &\leq 2\sqrt{18d} (32 \ln(d) \text{tr}[A^2] \|\pi\|_2^2)^{1/8} + \|\pi\|_2^2 && \text{choosing } k \text{ optimally} \\ &\leq 14 (d^4 \ln(d) \text{tr}[A^2] \|\pi\|_2^2)^{1/8} + \|\pi\|_2^2 \end{aligned}$$

In the line where we choose  $k$  optimally, we are choosing  $k$  to balance the two terms, and then rounding up to the nearest integer, which incurs an extra error of up to  $\|\pi\|_2^2$ .

To prove Eq. (63), we choose  $z$  to minimize  $\text{tr}(H\sigma_z)$ , and then further choose a pure product state  $|\varphi\rangle$  from the support of  $\sigma_z$  to minimize  $\text{tr}(H\varphi)$ .  $\square$

### 5.3 Proof of the basic version of the theorem

Here we sketch the proof of Theorem 3, mostly by pointing out places in which the proof of Theorem 8 can be modified.

The first change is that  $C$  can be immediately chosen to minimize  $\gamma$ , achieving  $\gamma \leq \ln(d)/k$ . Next, we can set  $\lambda = D/n$  and have  $\eta = 0$ . Thus we obtain

$$\langle G, \Delta \rangle \leq \sqrt{\langle G, \Delta^2 \rangle} \leq \sqrt{\lambda \langle \pi \pi^T, \Delta^2 \rangle} \leq \sqrt{\frac{2n \ln(d)}{kD}}.$$

This classical trace distance is again multiplied by  $18d$  to obtain the quantum trace distance. Measuring  $k$  vertices ruins a  $2k/n$  fraction of the edges. The total error is then

$$18d \sqrt{\frac{2n \ln(d)}{kD}} + \frac{2k}{n}.$$

We then choose  $k$  to balance these terms, and round up the nearest integer, incurring a further error that is  $\leq 1/n$ . This yields error

$$\left( \frac{1296d^2 \ln(d)}{D} \right)^{1/3} + \frac{1}{n}.$$

We obtain the claimed bound by using  $n \geq D$  and assuming that the first term is  $\leq 2$ .

## 6 Proof of amplification results: Proposition 10 and Corollary 11

**Proposition 10 (restatement).** *For any constants  $c, \alpha, \beta > 0$ , it is NP-hard to determine whether a 2-CSP  $\mathcal{C}$ , with alphabet  $\Sigma$  on a constraint graph of degree  $D$ , is such that  $\text{unsat}(\mathcal{C}) = 0$  or  $\text{unsat}(\mathcal{C}) \geq c|\Sigma|^\alpha/D^\beta$ .*

*Proof.* By the PCP theorem it follows there is a constant  $\varepsilon_0 > 0$  such that it is NP-hard to decide whether a 3-SAT(5) formula  $\mathcal{F}$  in which each variable belongs to 5 clauses is satisfiable or if  $\text{unsat}(\mathcal{F}) \geq \varepsilon_0$ .

The first step in the proof is to map this problem to a label cover problem (which is a 2-CSP). Here we follow Ref. [67]. A label cover problem is specified by the following parameters:

1. a bipartite graph  $G(V \cup W, E)$ ;
2. labels  $[N]$  and  $[M]$  for the vertices in  $V$  and  $W$ , respectively;
3. a function  $\{\Pi_{v,w}\}_{(v,w) \in E}$  on every edge  $(v,w) \in E$  such that  $\Pi_{v,w} : [M] \rightarrow [N]$ .

A labelling  $l : V \rightarrow [N], W \rightarrow [M]$  covers the edge  $(v,w)$  if  $\Pi_{v,w}(l(w)) = l(v)$ . The goal of the problem is to find a labelling of the vertices that covers the maximum number of edges.

The reduction from 3-SAT(5) with variables  $\{x_1, \dots, x_n\}$  and constraints  $\{C_1, \dots, C_m\}$  to label cover is as follows. Let  $V = \{x_1, \dots, x_n\}$  be the set of left vertices and  $W = \{C_1, \dots, C_m\}$  the set of right vertices. Two vertices are adjacent if, and only if,  $x_i \in C_j$ . The constraint  $\Pi_{x_i, C_j} : [7] \rightarrow [2]$  takes one of the possible 7 satisfying assignments for  $C_j$  and outputs the value of  $x_i$  in the assignment. Let's denote this label cover problem by  $\mathcal{L}$ . If  $\mathcal{F}$  is satisfiable then so is  $\mathcal{L}$ . On the other hand, if  $\text{unsat}(\mathcal{F}) \geq \varepsilon_0$ , then  $\text{unsat}(\mathcal{L}) \geq \varepsilon_0/3$  (see e.g. [67]). Therefore it is NP-hard to determine whether  $\text{unsat}(\mathcal{L}) = 0$  or  $\text{unsat}(\mathcal{L}) \geq \varepsilon_0/3$ .

Note that in  $\mathcal{L}$  the left vertices have degree 5 and the right vertices have degree 3, while the alphabet size  $|\Sigma| = 7$ . The next step is to increase the degree while making the graph regular. We proceed as follows: we create new variables  $\{\tilde{x}_i\}$  as left vertices and to each of the right vertices  $C_j(x_{i_1}, x_{i_2}, x_{i_3})$  we associate  $s = \lceil |\Sigma|^{2\alpha/\beta} \rceil - 3$  different  $\tilde{x}_i$  and add trivial constraints that are always satisfied between  $C_j$  and each of them, i.e. we replace  $C_j(x_{i_1}, x_{i_2}, x_{i_3})$  by  $C'_j(x_{i_1}, x_{i_2}, x_{i_3}, \tilde{x}_{k_1}, \dots, \tilde{x}_{k_s}) = C_j(x_{i_1}, x_{i_2}, x_{i_3})$ . In the process we also make sure that each of the new variables  $\tilde{x}_j$  is connected to  $|\Sigma|^{2\alpha/\beta}$  different right vertices. Then we create new constraints  $\tilde{C}_j$  and to each of the left vertices  $x_i$  we associate  $|\Sigma|^{2\alpha/\beta} - 5$  of them, while making sure that each of the new constraints  $\tilde{C}_j$  is also connected to  $\lceil |\Sigma|^{2\alpha/\beta} \rceil$  variables. Denote this new instance by  $\mathcal{C}$ . We perform this mapping to make sure that  $|\Sigma|^\alpha \leq \deg(\mathcal{C})^\beta$ , with  $\deg(\mathcal{C})$  the degree of  $\mathcal{C}$ . This procedure increases the total number of constraints by a multiplicative constant factor (depending on  $\alpha, \beta$ ). Therefore  $\text{unsat}(\mathcal{C}) \geq \Omega(\text{unsat}(\mathcal{L}))$ .

The next step is to apply parallel repetition to  $\mathcal{C}$ . For an integer  $t > 0$ , consider the following label cover instance  $\mathcal{C}_t$  specified by

1. a bipartite graph  $G'(V', W', E')$  with  $V' = V^t$  and  $W' = W^t$ ;
2. labels  $[N'] = [N]^t$  and  $[M'] = [M]^t$  for the vertices in  $V'$  and  $W'$ , respectively;
3. an edge set  $E'$  such that  $(v', w') \in E'$  if, and only if,  $(v_{i_j}, w_{i_j}) \in E$  for all  $i$  and  $1 \leq j \leq t$ , with  $v' = \{v_{i_1}, \dots, v_{i_t}\}$  and  $w' = \{w_{i_1}, \dots, w_{i_t}\}$ ;
4. a function  $\{\Pi_{v', w'}\}_{(v', w') \in E'}$  on every edge  $(v', w') \in E'$  such that  $\Pi_{v', w'} : [M'] \rightarrow [N']$  with  $\Pi_{v', w'}(b_1, \dots, b_t) = \{\Pi_{v_{i_1}, w_{i_1}}(b_1), \dots, \Pi_{v_{i_t}, w_{i_t}}(b_t)\}$ .

We claim that  $\text{unsat}(\mathcal{C}_t) \geq \text{unsat}(\mathcal{C})$ . Indeed one way to see this fact is by evoking Raz's parallel repetition theorem [64]. However here we only need the easy implication that the unsat is not decreased by parallel repetition (as opposed to the fact that the unsat approaches 1 exponentially in  $t$ ). Moreover the degree of each vertex is larger than  $|\Sigma|^{2t\alpha/\beta}$  and so  $\deg_t \geq |\Sigma|^{2t\alpha/\beta}$ , with  $\deg_t$  the average degree of  $\mathcal{C}_t$ , while the alphabet size of  $\mathcal{C}_t$  is  $|\Sigma_t| = |\Sigma|^t$ . Therefore for a constant  $t$  large enough  $\Omega(\varepsilon_0) \geq c|\Sigma|^{-t\alpha} \geq c|\Sigma_t|^\alpha / (\deg_t)^\beta$ .  $\square$

**Corollary 11 (restatement).** *Suppose that for every  $t \geq 1$  there is a mapping  $\mathcal{P}_t$  from 2-local Hamiltonians on  $d$ -dimensional particles to 2-local Hamiltonians on  $n_t$   $d_t$ -dimensional particles such that:*

- (i)  $\mathcal{P}_t$  can be computed deterministically in polynomial time.
- (ii)  $\mathcal{P}_t(H) = \mathbb{E}_{(i,j) \sim G'} H'_{i,j}$  with  $G'$  a probability distribution and  $\|H'_{i,j}\| \leq 1$ .
- (iii)  $\deg(\mathcal{P}_t(H)) \geq \deg(H)^t$ ,
- (iv)  $n_t \leq n^{O(t)}$ ,
- (v)  $d_t = d^t$ ,
- (vi)  $e_0(\mathcal{P}_t(H)) \geq e_0(H)$ , if  $e_0(H) > 0$ ,
- (vii)  $e_0(\mathcal{P}_t(H)) \leq e_0(H)$ , if  $e_0(H) \leq 0$ .

Then the quantum PCP conjecture is false.

*Proof.* Suppose we are given a 2-local Hamiltonian  $H$  with “PCP normalization”; i.e.  $H = \sum_{i,j} G_{i,j} H_{i,j}$ , with  $G$  a probability distribution (with  $G_{i,i} = 0$  for each  $i$ ) and  $H_{i,j}$  acting on qudits  $i, j$  with  $\|H_{i,j}\| \leq 1$ . Consider the problem of determining whether  $e_0(H) \leq 0$  or  $e_0(H) \geq \epsilon$  subject to the promise that one of these holds. Pad  $H$  with enough dummy edges to bring the degree to  $D$ , for  $D$  a parameter to be determined later. For each of these “dummy edges” the corresponding term in the Hamiltonian is  $H_{i,j} = 0$ . Let  $H'$  be the resulting Hamiltonian. Note that  $H' = cH$  for  $1/D \leq c \leq 1$ . Thus our two cases correspond to  $e_0(H') \leq 0$  or  $\geq \epsilon/D$ .

Now apply  $\mathcal{P}_t$  to  $H'$ . By assumption, we have either  $e_0(\mathcal{P}_t(H')) \leq 0$  or  $e_0(\mathcal{P}_t(H')) \geq \epsilon/D$ . However, by Corollary 5, we can approximate  $e_0(\mathcal{P}_t(H'))$  to accuracy  $12(d^{2t} \ln(d^t)/D^t)^{1/3}$  with a NP witness of size at most  $n^{O(t)}$ . (For simplicity, we neglect the additional arithmetic errors from discretization.) Now choose  $D = 8d^3$  and use  $\ln(d) \leq d$ , so that this accuracy is  $\leq 12 \cdot 2^{-t}$ . If  $t \geq \log_2(12D/\epsilon) = \log_2(96d^3/\epsilon)$ , then this will be enough to distinguish our two cases. Since  $d$  and  $\epsilon$  are constants, the witness size and the computation time are both polynomial in the original size of the computation. Thus, our original energy estimation problem is in NP, which implies the quantum PCP theorems is false (assuming that  $\text{NP} \neq \text{QMA}$ .)  $\square$

## 7 Proof of Theorem 12

**Theorem 12 (restatement).** *For every  $\epsilon > 0$  there is a polynomial-time randomized algorithm such that, given a 2-local Hamiltonian  $H$  on a planar graph, computes a number  $x$  such that  $e_0(H) \leq x \leq e_0(H) + \epsilon$ .*

*Proof.* Consider a partition of the vertices of the graph into two sets:  $\mathcal{H}$  and  $\mathcal{L}$ ;  $\mathcal{H}$  contain all the vertices with degree larger than  $f(\epsilon) := (23328d^2 \ln(d)/\epsilon)^2$ , while  $\mathcal{L}$  contains the remaining vertices.

Following the proof of Theorem 6, let  $p^{X_1, \dots, X_n} := \Lambda^{\otimes n}(\psi_0^{Q_1, \dots, Q_n})$  be a probability distribution obtained by measuring all  $n$  subsystems of  $\psi_0$  with the informationally-complete measurement  $\Lambda$  given by Lemma 16. Then by Lemma 21 and Markov’s inequality, there is a  $k' \leq k$  such that with probability larger than  $2/3$  over the choice of vertices  $c_1, \dots, c'_k$ ,

$$\mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \neq i} I(X_i : X_j | X_{c_1}, \dots, X_{c'_k}) \leq \frac{3 \ln(d)}{k}. \quad (74)$$

Also by Markov’s inequality with probability larger than  $2/3$ ,

$$\frac{1}{k'} \sum_{j=1}^{k'} \deg(c_j) \leq 3\bar{D}, \quad (75)$$

with  $\bar{D}$  the average degree of the graph. Thus by the union bound we can choose  $c_1, \dots, c'_k$  such that both Eqs. (74) and (75) hold true.

Let  $\tau$  be the state resulting from measuring  $\psi_0$  only on systems  $Q_{c_1}, \dots, Q_{c'_k}$ , and replacing the measured state on those systems with an arbitrary state, such as maximally mixed state. We can

write  $\tau = \mathbb{E}_{z \sim p^Z} \tau_z$ , where  $\tau_z$  is the state conditioned on obtaining measurement outcome  $z$ . Now define  $\sigma = \mathbb{E}_{z \sim p^Z} \sigma_z^{Q_1, \dots, Q_n}$ , with

$$\sigma_z^{Q_1, \dots, Q_n} := \left( \bigotimes_{i \in H} \tau_z^{Q_i} \right) \otimes \tau_z^{Q_L}, \quad (76)$$

where  $Q_L := \cup_{i \in L} Q_i$ ; i.e. we break the entanglement of the sites in the high-degree region  $\mathcal{H}$  and keep the original density matrix in the low-degree region  $\mathcal{L}$ . Then

$$e_0(H) \geq \text{tr}(H\tau) - \frac{6k}{n} \geq \text{tr}(H\sigma) - \mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{(i,j) \sim G} \|\tau^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 - \frac{6k}{n}, \quad (77)$$

where the first inequality follows from the fact that we measure  $k' \leq k$  subsystems and their average degree is at most triple the average degree (Eq. (75)), and the second from the fact that  $H$  is a 2-local Hamiltonian and that we consider a product ansatz only for the vertices in  $\mathcal{H}$ . Now define  $\Gamma(i)$  to the neighbors of  $i$  and use in turn the triangle inequality, Lemma 16, Pinsker's inequality and convexity to bound

$$\begin{aligned} \mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \sim \Gamma(i)} \|\tau^{Q_i Q_j} - \sigma^{Q_i Q_j}\|_1 &\leq \mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \sim \Gamma(i)} \mathbb{E}_{z \sim p^Z} \|\tau_z^{Q_i Q_j} - \sigma_z^{Q_i Q_j}\|_1 \\ &\leq 18d \mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \sim \Gamma(i)} \sqrt{2I(X_i : X_j | X_{c_1}, \dots, X_{c'_k})} \\ &\leq 18d \sqrt{2 \mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \sim \Gamma(i)} I(X_i : X_j | X_{c_1}, \dots, X_{c'_k})} \end{aligned} \quad (78)$$

Since  $\text{degree}(i) \geq f(\varepsilon)$  for all  $i \in \mathcal{H}$ , choosing  $k = \delta n$  and using Eq. (74) gives

$$\mathbb{E}_{i \in \mathcal{H}} \mathbb{E}_{j \sim \Gamma(i)} I(X_i : X_j | X_{c_1}, \dots, X_{c'_k}) \leq \frac{3 \ln(d)}{\delta f(\varepsilon)}. \quad (79)$$

Thus by Eq. (77)

$$\text{tr}(H\sigma) - e_0(H) \leq 2\delta + 18d \sqrt{\frac{6 \ln(d)}{\delta f(\varepsilon)}}. \quad (80)$$

Choosing  $\delta$  to balance the terms we get

$$\text{tr}(H\sigma) - e_0(H) \leq 18d \left( \frac{6 \ln(d)}{f(\varepsilon)} \right)^{1/2}, \quad (81)$$

and so by Eq. (76) there is a product state  $\pi := \bigotimes_{i \in H} \pi_i \otimes \pi_L$  such that

$$\text{tr}(H\pi) - e_0(H) \leq 18d \left( \frac{6 \ln(d)}{f(\varepsilon)} \right)^{1/2}. \quad (82)$$

Let us now focus on  $\pi_L$ . Consider the subgraph of  $G$  whose edges only act on the region  $\mathcal{L}$ . This is again a planar graph. By section 3.1 of [13] we can partition this subgraph into disjoint graphs each of size at most  $f(\varepsilon)^{O(1/\varepsilon)}$  by deleting at most  $\varepsilon n/3$  edges. Let  $\mathcal{L} = \{\mathcal{L}_1, \dots, \mathcal{L}_u\}$  be a partition of  $\mathcal{L}$  such that each  $\mathcal{L}_j$  contains the vertices of one of the disjoint subgraphs. Then

$$\text{tr} \left( H \bigotimes_{i \in \mathcal{H}} \pi_i \otimes \bigotimes_{j \in [u]} \sigma_{\mathcal{L}_j} \right) \leq \text{tr}(H\pi) + 18d \left( \frac{6 \ln(d)}{f(\varepsilon)} \right)^{1/2} + \varepsilon/3 \leq \text{tr}(H\pi) + \varepsilon/2 \quad (83)$$

Therefore it suffices to estimate

$$\max_{\psi_{\mathcal{H}_1}, \dots, \psi_{\mathcal{H}_k}, \psi_{\mathcal{L}_1}, \dots, \psi_{\mathcal{L}_u}} \langle \psi_{\mathcal{H}_1}, \dots, \psi_{\mathcal{H}_k}, \psi_{\mathcal{L}_1}, \dots, \psi_{\mathcal{L}_u} | H | \psi_{\mathcal{H}_1}, \dots, \psi_{\mathcal{H}_k}, \psi_{\mathcal{L}_1}, \dots, \psi_{\mathcal{L}_u} \rangle \quad (84)$$

to within error  $\varepsilon/2$ .

For this goal for each  $i \in \mathcal{H}$  and  $\mathcal{L}_j$  we consider a  $\delta$ -net on  $\mathbb{C}^{\tilde{d}}$ , with  $\tilde{d}$  the dimension of the space on  $i$  or  $\mathcal{L}_j$ . It is given by a set  $\mathcal{T}_\delta$  of less than  $(5/\delta)^{2\tilde{d}} \leq \exp(O(\ln(1/\delta) \exp(O(f(\varepsilon)^{O(1/\varepsilon)}))))$  states  $|\phi_k\rangle$  such that for all  $|\psi\rangle \in \mathbb{C}^{\tilde{d}}$  there is a  $k$  such that  $\|\psi - \phi_k\|_1 \leq \delta$ . We have

$$\min_{\{\psi_i\}_{i \in \mathcal{H}}, \{\psi_{\mathcal{L}_j}\}_{j \in [u]}} \text{tr}(H(\bigotimes_{i \in \mathcal{H}} \psi_i \otimes \bigotimes_{j \in [u]} \psi_{\mathcal{L}_j})) \leq \min_{\{\psi_i \in \mathcal{T}_\delta\}_{i \in \mathcal{H}}, \{\psi_{\mathcal{L}_j} \in \mathcal{T}_\delta\}_{j \in [u]}} \text{tr}(H(\bigotimes_{i \in \mathcal{H}} \psi_i \otimes \bigotimes_{j \in [u]} \psi_{\mathcal{L}_j})) - 2\delta \quad (85)$$

Therefore we can solve the problem by performing the optimization only over  $\mathcal{T}_\delta$  for a sufficiently small  $\delta$ . Under this restriction the problem becomes the minimization of the ground-state energy of a classical Hamiltonian (with  $|\mathcal{T}_\delta|$  possible values for the spin in each site) over a planar graph and it follows from [13] that one can obtain a  $\nu$ -additive approximation for this problem in polynomial-time for all  $\nu > 0$ . Finally we take  $\delta = \varepsilon/6$  and  $\nu = \varepsilon/6$ , so that our total error is  $\leq \varepsilon$ .  $\square$

## 8 Proof of Theorem 13

The next lemma is an easy corollary of a result of Gharibian and Kempe [39], who used it to give a polynomial-time algorithm achieving a  $d^{-k+1}$ -approximation for the problem of computing the maximum eigenvalue of dense  $k$ -local Hamiltonians on  $d$ -dimensional particles.

**Lemma 23.** *For all  $\varepsilon > 0$  there is a polynomial-time algorithm for dense  $k$ -local Hamiltonian  $H$  that outputs a product state assignment with value at most  $\varepsilon$  larger than*

$$\min_{\psi_1, \dots, \psi_n} \langle \psi_1, \dots, \psi_n | H | \psi_1, \dots, \psi_n \rangle. \quad (86)$$

*Proof.* Given a Hamiltonian  $H' = l^{-1} \sum_{j=1}^l Q_j$  with  $0 \leq Q_j \leq \text{id}$ , Theorem 4 of Ref. [39] gives an algorithm for computing a  $(1 + \varepsilon)$ -multiplicative approximation of  $\max_{|\psi_1, \dots, \psi_n\rangle} \langle \psi_1, \dots, \psi_n | H' | \psi_1, \dots, \psi_n \rangle$ . Let  $H = l^{-1} \sum_{j=1}^l H_j$ , with  $\|H_j\| \leq 1$ , and define  $H' = I - l^{-1} \sum_{j=1}^l H_j = l^{-1} \sum_{j=1}^l (I - H_j)$ . As  $\max_{|\psi_1, \dots, \psi_n\rangle} \langle \psi_1, \dots, \psi_n | H | \psi_1, \dots, \psi_n \rangle = 1 - \max_{|\psi_1, \dots, \psi_n\rangle} \langle \psi_1, \dots, \psi_n | H' | \psi_1, \dots, \psi_n \rangle$ , we find that we can estimate the L.H.S. within additive error  $\varepsilon$  using the  $(1 + \varepsilon)$ -multiplicative approximation of the second term of the R.H.S..  $\square$

We now turn to the main result of this section.

**Theorem 13 (restatement).** *For every  $\varepsilon > 0$  there is a polynomial-time algorithm such that, given a dense  $k$ -local Hamiltonian  $H$ , computes a number  $x$  such that  $e_0(H) \leq x \leq e_0(H) + \varepsilon$ .*

*Proof.* We will prove that there is a product state  $|\psi_1\rangle \otimes \dots \otimes |\psi_n\rangle$  such that

$$\langle \psi_1, \dots, \psi_n | H | \psi_1, \dots, \psi_n \rangle \leq e_0(H) + O(n^{-1/4}). \quad (87)$$



The result then follows from Lemma 23.

To prove Eq. (87) we use Theorem 17. Let  $\rho := |\psi_0\rangle\langle\psi_0|$  be the projector onto a ground state of  $|\psi_0\rangle$ . Then, given an integer  $t$ , we find that there are sites  $j_1, \dots, j_m$  such that

$$\begin{aligned} & \mathbb{E}_{a_1, \dots, a_m} \mathbb{E}_{i_1, \dots, i_k} \left\| \rho_{X_{j_1}=a_1, \dots, X_{j_m}=a_m}^{Q_{i_1}, \dots, Q_{i_k}} - \rho_{X_{j_1}=a_1, \dots, X_{j_m}=a_m}^{Q_{i_1}} \otimes \dots \otimes \rho_{X_{j_1}=a_1, \dots, X_{j_m}=a_m}^{Q_{i_k}} \right\|_1 \\ & \leq \sqrt{\frac{4 \ln(d)(18d)^k k^2}{t}}, \end{aligned} \quad (88)$$

where we used the convexity of  $x \mapsto x^2$ . Define

$$\pi := \Lambda_{Q_{j_1}} \otimes \dots \otimes \Lambda_{Q_{j_t}}(\rho) = \sum_{a_1, \dots, a_t} p(a_1, \dots, a_t) |a_1, \dots, a_t\rangle\langle a_1, \dots, a_t| \otimes \pi_{a_1, \dots, a_t}, \quad (89)$$

with  $p(a_1, \dots, a_t) := \text{tr}(M_{a_1} \otimes \dots \otimes M_{a_t} \rho_{Q_{j_1}, \dots, Q_{j_t}})$  and  $\pi_{a_1, \dots, a_t}$  the post-selected state on the remaining sites after measuring sites  $Q_{j_1}, \dots, Q_{j_m}$  of  $\rho$  and obtaining outcomes  $a_1, \dots, a_t$ . Also define the fully separable state

$$\sigma := \sum_{a_1, \dots, a_t} p(a_1, \dots, a_t) |a_1, \dots, a_t\rangle\langle a_1, \dots, a_t| \otimes \pi_{a_1, \dots, a_t}^{Q_{i_1}} \otimes \dots \otimes \pi_{a_1, \dots, a_t}^{Q_{i_l}}, \quad (90)$$

with  $\{i_1, \dots, i_l\} = [n] \setminus \{j_1, \dots, j_m\}$ . On one hand we have

$$\text{tr}(H\pi) \leq e_0(H) + \frac{t}{n}. \quad (91)$$

On the other hand by Eq. (88),

$$\mathbb{E}_{i_1, \dots, i_k} \left\| \pi^{Q_{i_1}, \dots, Q_{i_k}} - \sigma^{Q_{i_1}, \dots, Q_{i_k}} \right\|_1 \leq \sqrt{\frac{4 \ln(d)(18d)^k k^2}{t}} + \frac{t}{n} \quad (92)$$

Thus

$$\text{tr}(H\sigma) \leq \text{tr}(H\pi) + \sqrt{\frac{4 \ln(d)(18d)^k k^2}{t}} \leq e_0(H) + 2\frac{t}{n} + \sqrt{\frac{4 \ln(d)(18d)^k k^2}{t}}. \quad (93)$$

Choosing  $t = \sqrt{n}$  gives the result.  $\square$

## 9 SDP hierarchies for the ground-state energy on graphs with low threshold rank

In this section, we analyze the use of the Lasserre hierarchy for lower-bounding the ground-state energy of a Hamiltonian and prove Theorem 14.

*Definition of the hierarchy:* Given a Hamiltonian  $H \in W_k$  (in fact, we will consider here only  $H \in W_2$ ), we will minimize  $M(H, I)$  over bilinear functions  $M : W_k \times W_k \rightarrow \mathbb{R}$  satisfying

$$M(I, I) = 1 \quad (94a)$$

$$M(X, Y) = M(X', Y') \quad \text{whenever } XY = X'Y' \quad (94b)$$

$$M(X, X) \geq 0 \quad \forall X \in W_k \quad (94c)$$

Of course  $M$  can be represented as a matrix of dimension  $\leq \binom{n}{k} d^{2k}$ , and the optimization over  $M$  can be performed by semidefinite programming, since it suffices to check Eq. (94b) on a basis for  $W_k$  and Eq. (94c) is equivalent to the condition that  $M \succeq 0$ .

To see that this is indeed a relaxation of the original problem, we observe that if  $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$  then  $M(X, Y) := \text{tr}(\rho XY)$  is a valid solution. Clearly it satisfies the normalization constraint (94a) and the constraint (94b) that  $M(X, Y)$  depends only on  $X, Y$ . For the PSD constraint (94c), observe that for any  $X \in W_k$  we have  $M(X, X) = \text{tr}(\rho X^2) \geq 0$ , since  $\rho, X^2$  are both PSD.

Since  $M(X, Y)$  depends only on  $XY$ , we abuse notation and write  $M(XY) := M(X, Y)$ . Indeed, an equivalent formulation of the hierarchy is that we specify all reduced density matrices of  $\leq 2k$  systems. However, in this formulation, the condition that  $M$  is PSD is less transparent. It will turn out that this global PSD constraint is crucial. Classically a similar situation holds when this hierarchy (there known as the Lasserre hierarchy) is applied to combinatorial optimization problems. In this case, dropping the global PSD condition results in a linear program (the Sherali-Adams relaxation) expressing only the condition that the marginal distributions be compatible. (In the quantum case, semidefinite programming is still required even to properly define the marginals.) It is known that this weaker form of optimization requires  $k = \Omega(n)$  to give good approximations of many CSPs [25, 18]; and indeed such negative results are also known for the Lasserre hierarchy [68]. On the other hand, in [14] the Lasserre hierarchy was shown to give good approximations of 2-CSPs on graphs with low threshold-rank. These are defined as graphs for which the threshold rank  $\text{rank}_\lambda$ , defined as the number of eigenvalues with value larger than  $\lambda$ , is bounded.

Theorem 14 will follow by an adaptation of the techniques of [14]. In fact our algorithm for estimating the ground-state energy is essentially the same as the Propagation Sampling algorithm (Algorithm 5.5) of [14]. The main new element is our introduction of state tomography.

Imagine that we measure each qudit with an informationally-complete POVM  $\Lambda$  with  $\tilde{d} := \text{poly}(d)$  outcomes, corresponding to measurement operators  $\{E_1, \dots, E_{\tilde{d}}\}$ . The outcomes of these POVM yield exactly the same information as  $M$ , and thus constitute an equivalent optimization. However, the outcomes of these measurements can be interpreted (using the language of [14]) as  $2k$ -local random variables with range  $[\tilde{d}]$ . This terminology means that for any  $i_1, \dots, i_{2k} \in [n]$  we can define a probability distribution  $p^{X_{i_1} \dots X_{i_{2k}}}$  and that the marginals of these distributions are consistent for overlapping sets. We let  $X_i$  (for  $i \in [n]$ ) denote the random variable corresponding to the measurement of qudit  $i$ , and for each  $x \in [\tilde{d}]$  let  $X_{i,x} = 1_{\{X_i=x\}}$  be the indicator R.V. corresponding to a particular outcome of the measurement. We will abuse notation and call  $p$  a probability distribution with the understanding that only its marginals on  $\leq 2k$  systems are actually defined. Since  $\Lambda$  is informationally complete,  $p$  carries exactly the same information as  $M$  does.

**Algorithm 24** (Quantum Propagation Sampling, based on Algorithm 5.5 of [14]).

**Input:** A matrix  $M$  that is a valid solution to Eq. (94).

**Output:** A product state  $\sigma = \sigma_1 \otimes \cdots \otimes \sigma_n$ .

1. Choose  $m \in \{0, \dots, 2k - 1\}$  at random.
2. Choose  $i_1, \dots, i_m \in [n]$  randomly with replacement, and set  $S = \{i_1, \dots, i_m\}$ .
3. Sample an outcome  $x_S = (x_{i_1}, \dots, x_{i_m})$  with probability  $p^{X_S}(x_S)$ .
4. Let  $p|_{X_S=x_S}$  be the  $2k - m$ -local distribution resulting from conditioning on the outcome  $X_S = x_S$ .
5. For every qudit  $i \in [n] \setminus S$ , set  $\sigma_i$  to be the unique density matrix that fits the measurement outcomes  $p^{X_i}|_{X_S=x_S}$ ; i.e.  $\sigma_i = \Lambda^{-1}(p^{X_i}|_{X_S=x_S})$ .
6. For each  $i \in S$ , set  $\sigma_i$  arbitrarily.

**Theorem 14 (restatement).** Let  $G$  be a graph, and  $H = \mathbb{E}_{(i,j) \sim G} H_{i,j}$  a 2-local Hamiltonian on  $n$  qudits with each  $\|H_{i,j}\| \leq 1$ . Then it is possible to estimate the ground-state energy of  $H$  to within additive error  $\epsilon$  in time  $\exp(\text{poly}(d/\epsilon) \text{rank}_{\text{poly}(\epsilon/d)}(G)) \cdot \text{poly}(n)$ . If there exists a measurement  $\{A_x\}$  (i.e.  $A_x \geq 0$  and  $\sum_x A_x = I$ ) and a set of permutations  $\pi_{i,j}$  such that each  $H_{i,j} = -\sum_x A_x \otimes A_{\pi_{i,j}(x)}$  then the algorithm will run in time  $\exp(\log(d/\epsilon)) \cdot \text{poly}(n)$ .

*Proof.* We will prove that Algorithm 24 achieves the error bound claimed by Theorem 14. First, we need to argue that step 5 always produces legal density matrices. To see this, observe that  $M$  can be used to define an  $m + 1$ -partite density matrix  $\rho^{S \cup \{i\}}$  satisfying  $\text{tr}(\rho^{S \cup \{i\}} X) = M(X^{Q_S Q_i})$  for any operator  $X$  on  $(\mathbb{C}^d)^{\otimes m+1}$ . Then we see that  $p^{X_S}(x_S) = \text{tr}(\rho^{S \cup \{i\}}(E_{x_{i_1}}^{Q_{i_1}} \otimes \dots \otimes E_{x_{i_m}}^{Q_{i_m}} \otimes I))$  and  $\sigma_i = \text{tr}_S(\rho^{S \cup \{i\}}(E_{x_{i_1}}^{Q_{i_1}} \otimes \dots \otimes E_{x_{i_m}}^{Q_{i_m}} \otimes I))/p^{X_S}(x_S)$ . This is simply the state resulting from obtaining measurement outcomes  $x_{i_1}, \dots, x_{i_m}$  on systems  $i_1, \dots, i_m$ , and thus must be a valid density matrix.

Now we argue that  $\text{tr}(H\sigma) \approx M(H)$ . This will use the analysis of a closely related algorithm in [14]. Theorem 5.6 of [14] states that if  $p$  is the set of measurement outcomes induced from  $M$  and  $q$  is the product distribution on measurement outcomes produced by step 4 of Algorithm 24 (i.e.  $q = \bigotimes_i p^{X_i}|_{X_S=x_S}$ ), then under certain conditions, we have

$$\mathbb{E}_{(i,j) \sim G} \|p^{X_i X_j} - q^{X_i X_j}\|_1 \leq \delta. \quad (95)$$

Those conditions are that:

$$k \geq C \frac{\tilde{d}}{\delta^4} \text{rank}_{\Omega(\delta/\tilde{d})^2}(G) \quad (96)$$

$$(\text{Cov}_{p|_{X_S=x_S}}(X_{iy}, X_{jz}))_{iy,jz} \text{ is PSD for any } S, x_S \quad (97)$$

Satisfying (96) will simply require taking  $k$  large enough. For Eq. (97), we consider an arbitrary

vector  $v \in \mathbb{R}^{(n-m)\tilde{d}}$ , define  $Z = \sum_{i \in [n] \setminus S} \sum_{y \in [\tilde{d}]} v_{i,y} E_y^{X_i}$  and observe that

$$\sum_{i,j \in [n] \setminus S} \sum_{y,z \in [\tilde{d}]} v_{i,y} v_{j,z} \text{Cov}_{p|_{X_S=x_S}}(X_{iy}, X_{jz}) \quad (98)$$

$$= \sum_{i,j \in [n] \setminus S} \sum_{y,z \in [\tilde{d}]} v_{i,y} v_{j,z} \frac{M(E_{x_S}^{X_S} E_y^{X_i} E_z^{X_j}) M(E_{x_S}^{X_S}) - M(E_{x_S}^{X_S} E_y^{X_i}) M(E_{x_S}^{X_S} E_z^{X_j})}{M(E_{x_S}^{X_S})^2} \quad (99)$$

$$= \frac{M(Z^2 E_{x_S}^{X_S}) M(E_{x_S}^{X_S}) - M(Z E_{x_S}^{X_S})^2}{M(E_{x_S}^{X_S})^2} \quad (100)$$

$$= M'(Z^2) - M'(Z)^2. \quad (101)$$

To obtain this final simplification, we define a functional  $M'$  that acts on  $2k - m$ -local operators on systems  $[n] \setminus S$  by

$$M'(Y) := \frac{M(E_{x_S}^{X_S} Y)}{M(E_{x_S}^{X_S})}. \quad (102)$$

We claim that  $M'$  satisfies (94). Eqs. (94a) and (94b) are straightforward to verify. For Eq. (94c), we use the fact that  $E_x$  are all rank-1 projectors and  $E_{x_S}^{X_S}$  and  $Y$  act on disjoint systems to obtain

$$M'(Y^2) = \frac{M((\sqrt{E_{x_S}^{X_S}} Y)^2)}{M(\sqrt{E_{x_S}^{X_S}})^2} \geq 0. \quad (103)$$

Finally we set  $Y = Z - M'(Z)I$  in Eq. (101) to complete the proof of Eq. (97). This establishes Eq. (95).

To relate this to quantum states, define  $\rho_{ij}$  to be the two-qudit density matrix satisfying  $\text{tr}(M X^{Q_i Q_j}) = \text{tr}(\rho_{i,j} X)$ . The objective value of the SDP is  $\mathbb{E}_{(i,j) \sim E} \text{tr}(\rho_{i,j} H_{i,j}) \leq e_0(H)$ , so we will complete our proof by arguing that  $\rho_{i,j} \approx \sigma_i \otimes \sigma_j$ . Observe that  $p^{X_i X_j} = \Lambda^{\otimes 2}(\rho_{i,j})$  and  $q^{X_i X_j} = \Lambda(\sigma_i) \otimes \Lambda(\sigma_j)$ . Thus  $\|\rho_{i,j} - \sigma_i \otimes \sigma_j\|_1 \leq 18d \|p^{X_i X_j} - q^{X_i X_j}\|_1$  and

$$\text{tr}(H\sigma) - M(H) = \mathbb{E}_{(i,j)} \text{tr}(H_{i,j}(\sigma_i \otimes \sigma_j - \rho_{i,j})) \quad (104)$$

$$\leq \mathbb{E}_{(i,j)} \|\sigma_i \otimes \sigma_j - \rho_{i,j}\|_1 \quad (105)$$

$$\leq 18d \mathbb{E}_{(i,j)} \|q^{X_i X_j} - p^{X_i X_j}\|_1 \quad (106)$$

$$\leq 18d\delta \quad (107)$$

We complete the proof by setting  $\delta = \epsilon/18d$ .

Finally, for the special case where  $H_{i,j}$  is of the form  $-\sum_x A_x \otimes A_{\pi_{i,j}(x)}$ , Ref. [14] shows that Eq. (96) can be replaced by the condition that  $k \geq \text{poly}(1/\epsilon)$ . In this case, we let  $p$  be the distribution resulting from measuring each system according to  $\{A_x\}$ ; i.e.  $p(X_{i_1} = x_{i_1}, \dots, X_{i_m} = x_{i_m}) = M(A_{x_{i_1}}^{Q_{i_1}} \otimes \dots \otimes A_{x_{i_m}}^{Q_{i_m}})$ . Then [14] shows not that  $\|p^{X_i X_j} - q^{X_i X_j}\|_1$  is small, but instead that  $\sum_x |p^{X_i X_j}(x, \pi_{i,j}(x)) - q^{X_i X_j}(x, \pi_{i,j}(x))|$  is small. Since this is all we need to bound  $\text{tr}(H\sigma) - M(H)$ , we obtain the desired bound.  $\square$

## 10 Proofs of de Finetti theorems without symmetry

**Theorem 15 (restatement).** Let  $p^{X_1, \dots, X_n}$  be a probability distribution on  $\Sigma^n$ , for a finite set  $\Sigma$ . Let  $\vec{i} = (i_1, \dots, i_k)$ ,  $\vec{j} = (j_1, \dots, j_m)$  be random disjoint ordered subsets of  $[n]$ , and let  $\vec{x} = (x_1, \dots, x_m)$  be distributed according to  $p^{X_{\vec{j}}} := p^{X_{j_1} \dots X_{j_m}}$ . Define  $p_{X_{\vec{j}}=\vec{x}}$  to be the conditional distribution resulting from taking  $X_{j_1} = x_1, \dots, X_{j_m} = x_m$ , and define  $p_{X_{\vec{j}}=\vec{x}}^{X_{\vec{i}}}$  to be the  $X_{i_1}, \dots, X_{i_k}$  marginal of this distribution.

Then for every integer  $t \leq n - k$  there is an integer  $m \leq t$  such that

$$\mathbb{E}_{\substack{\vec{j}=(j_1, \dots, j_m) \\ \vec{x}=(x_1, \dots, x_m)}} \mathbb{E}_{\vec{i}=(i_1, \dots, i_k)} \left\| p_{X_{\vec{j}}=\vec{x}}^{X_{\vec{i}}} - p_{X_{\vec{j}}=\vec{x}}^{X_{i_1}} \otimes \dots \otimes p_{X_{\vec{j}}=\vec{x}}^{X_{i_k}} \right\|_1^2 \leq \frac{2k^2 \ln |\Sigma|}{t}, \quad (108)$$

*Proof.* By the chain rule of mutual information (Eq. (25) of Lemma 18):

$$I(X_{i_1} \dots, X_{i_k} : X_{j_1}, \dots, X_{j_t}) = I(X_{i_1} \dots, X_{i_k} : X_{j_1}) + \dots + I(X_{i_1} \dots, X_{i_k} : X_{j_t} | X_{j_1} \dots X_{j_{t-1}}). \quad (109)$$

Taking the expectation over  $i_1, \dots, i_k, j_1, \dots, j_t$ :

$$\begin{aligned} & \mathbb{E}_{j_1, \dots, j_t} \mathbb{E}_{i_1, \dots, i_k} I(X_{i_1} \dots, X_{i_k} : X_{j_1}, \dots, X_{j_t}) \\ &= \mathbb{E}_{j_1, \dots, j_t} \mathbb{E}_{i_1, \dots, i_k} I(X_{i_1} \dots, X_{i_k} : X_{j_1}) + \dots + \mathbb{E}_{j_1, \dots, j_t} \mathbb{E}_{i_1, \dots, i_k} I(X_{i_1} \dots, X_{i_k} : X_{j_t} | X_{j_1} \dots X_{j_{t-1}}). \end{aligned} \quad (110)$$

Thus there is a  $m \leq t$  such that (relabeling  $j_m$  to  $i_{k+1}$ )

$$\begin{aligned} & \mathbb{E}_{j_1, \dots, j_{m-1}} \mathbb{E}_{i_1, \dots, i_{k+1}} I(X_{i_1} \dots, X_{i_k} : X_{i_{k+1}} | X_{j_1} \dots X_{j_{m-1}}) \\ & \leq \frac{1}{t} \mathbb{E}_{j_1, \dots, j_t} \mathbb{E}_{i_1, \dots, i_k} I(X_{i_1} \dots, X_{i_k} : X_{j_1}, \dots, X_{j_t}) \\ & \leq \frac{k \log |\Sigma|}{t}, \end{aligned} \quad (111)$$

where the last inequality follows from Eq. (30).

Finally, from Eqs. (26) and (27):

$$\begin{aligned} & \mathbb{E}_{j_1, \dots, j_{m-1}} \mathbb{E}_{i_1, \dots, i_{k+1}} I(X_{i_1} : \dots : X_{i_{k+1}} | X_{j_1} \dots, X_{j_{m-1}}) \\ &= \mathbb{E}_{j_1, \dots, j_{m-1}} \mathbb{E}_{i_1, \dots, i_k} (I(X_{i_1} : X_{i_2} | X_{j_1} \dots, X_{j_{m-1}}) + \dots + I(X_{i_1} \dots X_{i_k} : X_{i_{k+1}} | X_{j_1} \dots, X_{j_{m-1}})) \\ & \leq k \mathbb{E}_{j_1, \dots, j_{m-1}} \mathbb{E}_{i_1, \dots, i_{k+1}} I(X_{i_1} \dots X_{i_k} : X_{i_{k+1}} | X_{j_1} \dots, X_{j_{m-1}}) \\ & \leq \frac{k^2 \ln |\Sigma|}{t}. \end{aligned} \quad (112)$$

The statement then follows by applying Pinsker's inequality (Eq. (29) of Lemma 18) to the quantity  $I(X_{i_1} : \dots : X_{i_k} | X_{j_1} \dots, X_{j_{m-1}})$ .  $\square$

## 11 Open Problems

There are several open problems related to the results of this paper.

- We start listing open problems mentioned before in the paper: To prove conjecture 9 in section 2.1.3, to improve the dependence on  $k$  in Theorem 17 from  $d^k$  to  $\text{poly}(k)$ , to obtain a generalization of Theorem 17 for the local norm (see discussion in section 2.3), and to obtain a generalization of Theorem 6 for non-regular graphs.
- An important open question is whether the quantum PCP conjecture holds true. We believe the results of this paper put into check the validity of the conjecture. At least they show that a quantum analogue of the PCP theorem in conjunction with parallel repetition of 2-CSPs (in order to increase the degree of the constraint graph, but without the need to amplify the gap as in Raz's theorem [64]; see the proof of Proposition 10) is *false*. Hence it emerges as an interesting problem whether there is a meaningful notion of parallel repetition for quantum CSPs. Note that parallel repetition of CSPs involves massive copying of the variables. This is the reason why we do not know how to generalize it to the quantum case, where the assignment is a highly entangled state and thus there is no clear notion of copying of the variables.
- There has been some recent activity in the complexity of the local Hamiltonian problem for commuting Hamiltonians [43, 23, 69, 4, 42]. On one hand it is possible that the problem is always in NP, even for estimating the energy to inverse polynomial accuracy. On the other hand, it is possible that it is QMA-hard to obtain a constant error approximation to the ground-state energy, settling the quantum PCP conjecture in the affirmative. For 2-local commuting Hamiltonians Bravyi and Vyalii proved that the problem is always in NP by  $C^*$ -algebraic techniques [23]. Later these techniques were generalized to more general Hamiltonians in [69, 4, 43]. Can this approach be combined with the methods of this paper to give a disproof of the quantum PCP conjecture for commuting Hamiltonians? One challenge would be to develop similar techniques for  $k$ -local Hamiltonians with  $k \geq 3$ , since this is the relevant case for commuting Hamiltonians<sup>10</sup>.
- Can we get in NP an approximation to the ground energy within small error of general  $k$ -local Hamiltonians on very good small-set expander graphs? This would be a common generalization of Theorem 8 (which establishes that for 2-local models) and Aharonov and Eldar result [4] for commuting  $k$ -local models.
- We have mentioned that our work is in the spirit of the mean-field approximation, and can be seen as a generalization of mean-field to Hamiltonians where different edges carry different interactions. Mean-field theory, however, is more than merely approximating states with product states. For example, there is a well-understood theory of corrections to mean-field theory, in which a tensor power state is viewed as a vacuum and fluctuations are viewed as bosonic excitations (see e.g. [66]). It would be interesting to have a systematic method of computing corrections to our product-state approximation for high-degree graphs.

---

<sup>10</sup>While for addressing the general quantum PCP conjecture it suffices to consider 2-local models (since one can apply perturbation theory to reduce the general case to 2-local Hamiltonians [22]), this is not the true for commuting Hamiltonians, since we do not have perturbation-theory gadgets that preserve commutativity.

- Finally it would be interesting to develop information-theoretic tools to analyse the performance of variational ansatz beyond product states. For example, what is the power of constant-depth circuits applied to product states (see [36])? Can information-theoretical methods give new results in this direction?

## Acknowledgements

We thank Dorit Aharonov, Itai Arad, Ignacio Cirac, Sev Gharibian, Will Matthews, David Poulin and Or Sattath for interesting discussions and useful comments. Much of this work was done while FGSLB was working at the Institute for Theoretical Physics in ETH Zürich and AWH was working in the Department of Computer Science at the University of Washington. FGSLB acknowledges support from EPSRC, the Polish Ministry of Science and Higher Education Grant No. IdP2011 000361, the Swiss National Science Foundation, via the National Centre of Competence in Research QSIT, the German Science Foundation (grant CH 843/2-1), the Swiss National Science Foundation (grants PP00P2\_128455, 20CH21\_138799 (CHIST-ERA project CQC)), the Swiss National Center of Competence in Research “Quantum Science and Technology (QSIT)”, and the Swiss State Secretariat for Education and Research supporting COST action MP1006. AWH was funded by NSF grants CCF-0916400 and CCF-1111382 and ARO contract W911NF-12-1-0486.

## References

- [1] S. Aaronson. The quantum PCP manifesto, 2006. <http://www.scottaaronson.com/blog/?p=139>.
- [2] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The detectability lemma and quantum gap amplification. In *STOC '09*, pages 417–426, 2009, arXiv:0811.3412.
- [3] D. Aharonov, I. Arad, and T. Vidick. Guest column: the quantum PCP conjecture. *SIGACT News*, 44(2):47–49, 2013.
- [4] D. Aharonov and L. Eldar. On the complexity of commuting local Hamiltonians, and tight conditions for topological order in such systems. In *FOCS '11*, pages 334–343, 2011, arXiv:1102.0770.
- [5] D. Aharonov and L. Eldar. Commuting local Hamiltonians on expanders, locally testable quantum codes, and the qPCP conjecture, 2013, arXiv:1301.3407.
- [6] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Commun. Math. Phys.*, 287(1):41–65, 2009, arXiv:0705.4077.
- [7] N. Alon, W. De La Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of MAX-CSP problems. In *STOC '02*, pages 232–239, 2002.
- [8] I. Arad. A note about a partial no-go theorem for quantum PCP. *Quant. Inf. Comp.*, 11(11-12):1019–1027, 2011, arXiv:1012.3319.
- [9] S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. In *STOC '95*, pages 284–293, 1995.

- [10] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *JACM*, 45(3):501–555, 1998.
- [11] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *JACM*, 45(1):70–122, 1998.
- [12] B. Baker. Approximation algorithms for NP-complete problems on planar graphs. *JACM*, 41(1):153–180, 1994.
- [13] N. Bansal, S. Bravyi, and B. M. Terhal. Classical approximation schemes for the ground-state energy of quantum and classical Ising spin Hamiltonians on planar graphs. *Quant. Inf. Comp.*, 9(7):701–720, 2009, arXiv:0705.1115.
- [14] B. Barak, P. Raghavendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS '11*, 2011, arXiv:1104.4680.
- [15] T. Barthel and R. Hübener. Solving condensed-matter ground-state problems by semidefinite relaxations. *Physical Review Letters*, 108(20):200404, 2012, arXiv:1106.4966.
- [16] T. Baumgratz and M. Plenio. Lower bounds for ground states of condensed matter systems. *New Journal of Physics*, 14(2):023027, 2012, arXiv:1106.5275.
- [17] C. Bazgan, W. Fernandez de la Vega, and M. Karpinski. Polynomial time approximation schemes for dense instances of minimum constraint satisfaction. *Random Structures & Algorithms*, 23(1):73–91, 2003.
- [18] S. Benabbas, K. Georgiou, A. Magen, and M. Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(12):269–289, 2012.
- [19] A. D. Bookatz. QMA-complete problems, 2012, arXiv:1212.6312.
- [20] F. G. Brandao and A. W. Harrow. Quantum de Finetti theorems under local measurements with applications. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, STOC '13, pages 861–870, 2013, arXiv:1210.6367.
- [21] F. G. S. L. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Commun. Math. Phys.*, 306(3):805–830, 2011, arXiv:1010.1750.
- [22] S. Bravyi, D. DiVincenzo, D. Loss, and B. Terhal. Quantum simulation of many-body Hamiltonians using perturbation theory with bounded-strength interactions. *Phys. Rev. Lett.*, 101(7):70503, 2008, arXiv:0803.2686.
- [23] S. Bravyi and M. Vyalyi. Commutative version of the local Hamiltonian problem and common eigenspace problem. *Quant. Inf. Comp.*, 5(3):187–215, May 2005, arXiv:quant-ph/0308021.
- [24] C. M. Caves, C. A. Fuchs, and R. Schack. Unknown quantum states: The quantum de Finetti representation. *J. Math. Phys.*, 43(9):4537–4559, 2002, arXiv:quant-ph/0104088.
- [25] M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 283–292, 2009.



- [26] M. Christandl, R. König, G. Mitchison, and R. Renner. One-and-a-half quantum de Finetti theorems. *Commun. Math. Phys.*, 273:473–498, 2007, arXiv:quant-ph/0602130.
- [27] W. F. de la Vega, M. Karpinski, R. Kannan, and S. Vempala. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *STOC '05*, pages 747–754, 2005.
- [28] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Annals of Probability*, 8:745–764, 1980.
- [29] I. Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3), June 2007.
- [30] A. Doherty, Y. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *CCC '08*, pages 199–210, 2008, arXiv:0803.4373.
- [31] A. C. Doherty and S. Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere, 2012, arXiv:1210.5048.
- [32] J. Eisert, M. Cramer, and M. Plenio. Area laws for the entanglement entropy - a review. *Reviews of Modern Physics*, 82(1):277, 2010, arXiv:0808.3773.
- [33] R. M. Erdahl. Representability. *International Journal of Quantum Chemistry*, 13(6):697–718, 1978.
- [34] M. Fannes, B. Nachtergaele, and R. Werner. Finitely correlated states on quantum spin chains. *Commun. Math. Phys.*, 144:443–490, 1992.
- [35] W. Fernandez De La Vega. MAX-CUT has a randomized approximation scheme in dense graphs. *Random Structures & Algorithms*, 8(3):187–198, 1996.
- [36] M. H. Freedman and M. B. Hastings. Quantum systems on non- $k$ -hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs, 2013, arXiv:1301.1363.
- [37] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *FOCS '96*, pages 12–20, 1996.
- [38] S. Gharibian and J. Kempe, 2012. personal communication.
- [39] S. Gharibian and J. Kempe. Approximation algorithms for QMA-complete problems. *SIAM J. Comput.*, 41(4):1028–1050, 2012, arXiv:1101.3884.
- [40] D. Gottesman and S. Irani. The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems. In *FOCS '09*, pages 95–104, 2009, arXiv:0905.2419.
- [41] A. W. Harrow. The church of the symmetric subspace, 2013, arXiv:1308.6595.
- [42] M. Hastings. Matrix product operators and central elements: Classical description of a quantum state, 2012, arXiv:1207.1671.
- [43] M. B. Hastings. Trivial low energy states for commuting Hamiltonians, and the quantum PCP conjecture, 2012, arXiv:1201.3387.
- [44] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Commun. Math. Phys.*, 246:359, 2004, arXiv:quant-ph/0304007.

- [45] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–562, 2006.
- [46] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009, [arXiv:quant-ph/0702225](#).
- [47] R. L. Hudson and G. R. Moody. Locally normal symmetric states and an analogue of de Finetti’s theorem. *Z. Wahrschein. verw. Geb.*, 33:343–351, 1976.
- [48] B. Ibinson, N. Linden, and A. Winter. Robustness of quantum Markov chains. *Commun. Math. Phys.*, 277(2):289–304, 2008, [arXiv:quant-ph/0611057](#).
- [49] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local Hamiltonian problem. *SIAM J. Comput.*, 35(5):1070–1097, 2006, [arXiv:quant-ph/0406180](#).
- [50] A. Y. Kitaev, A. H. Shen, and M. N. Vyalı. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. AMS, 2002.
- [51] R. Koenig and R. Renner. A de Finetti representation for finite symmetric quantum states. *J. Math. Phys.*, 46(12):122108, 2005, [arXiv:quant-ph/0410229](#).
- [52] B. Kraus, M. Lewenstein, and J. Cirac. Ground states of fermionic lattice Hamiltonians with permutation symmetry. 2013, [arXiv:arXiv:1305.4577](#).
- [53] C. Lancien and A. Winter. Distinguishing multi-partite states by local measurements, 2012, [arXiv:1206.2884](#).
- [54] D. A. Mazziotti. Realization of quantum chemistry without wave functions through first-order semidefinite programming. *Phys. Rev. Lett.*, 93:213001, Nov 2004.
- [55] A. Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12):122206, 2012, [arXiv:1208.0161](#).
- [56] H. Nakatsuji and K. Yasuda. Direct determination of the quantum-mechanical density matrix using the density equation. *Phys. Rev. Lett.*, 76:1039–1042, Feb 1996.
- [57] M. Navascues, A. Garcia-Saez, A. Acin, S. Pironio, and M. Plenio. A paradox in bosonic energy calculations via semidefinite programming relaxations, 2012, [arXiv:1203.3777](#).
- [58] R. Oliveira and B. M. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Info. Comput.*, 8(10):900–924, Nov. 2008, [arXiv:quant-ph/0504050](#).
- [59] T. Osborne. Hamiltonian complexity, 2011, [arXiv:1106.5875](#).
- [60] S. Pironio, M. Navascues, and A. Acin. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization*, 20(5):2157–2180, 2010, [arXiv:0903.4368](#).
- [61] D. Poulin and M. Hastings. Markov entropy decomposition: a variational dual for quantum belief propagation. *Physical review letters*, 106(8):80403, 2011, [arXiv:1012.2050](#).

- [62] G. A. Raggio and R. F. Werner. Quantum statistical mechanics of general mean field systems. *Helv. Phys. Acta*, 62:980–1003, 1989.
- [63] P. Raghavendra and N. Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *SODA '12*, pages 373–387, 2012, arXiv:1110.1064.
- [64] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.
- [65] R. Renner. Symmetry implies independence. *Nature Physics*, 3:645–649, 2007, arXiv:quant-ph/0703069.
- [66] I. Rodnianski and B. Schlein. Quantum fluctuations and rate of convergence towards mean field dynamics, 2007, arXiv:0711.3087.
- [67] M. Salavatipour. Label cover, hardness of set cover, 2011. <http://webdocs.cs.ualberta.ca/~mreza/courses/Approx11/week12.pdf>.
- [68] G. Schoenebeck. Linear level Lasserre lower bounds for certain  $k$ -CSPs. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS '08*, pages 593–602, 2008.
- [69] N. Schuch. Complexity of commuting Hamiltonians on a square lattice of qubits. *Quantum Information & Computation*, 11(11-12):901–912, 2011, arXiv:1105.2843.
- [70] N. Schuch and F. Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5(10):732–735, 2009, arXiv:0712.0483.
- [71] N. Schuch, M. Wolf, F. Verstraete, and J. Cirac. Entropy scaling and simulability by matrix product states. *Physical review letters*, 100(3):30504, 2008, arXiv:0705.0292.
- [72] E. Størmer. Symmetric states of infinite tensor products of  $C^*$ -algebras. *J. Funct. Anal.*, 3:48, 1969.
- [73] B. Terhal. Is entanglement monogamous?, 2003, arXiv:quant-ph/0307120.
- [74] V. Vazirani. *Approximation Algorithms*. Springer, 2010.
- [75] F. Verstraete and J. Cirac. Matrix product states represent ground states faithfully. *Physical Review B*, 73(9):094423, 2006, arXiv:cond-mat/0505140.
- [76] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91(14):147902, Oct 2003, arXiv:quant-ph/0301063.
- [77] J. Watrous. Quantum computational complexity, 2008, arXiv:0804.3401.
- [78] R. F. Werner. An application of Bell’s inequalities to a quantum state extension problem. *Lett. Math. Phys.*, 17:359, 1989.
- [79] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac. Area laws in quantum systems: Mutual information and correlations. *Phys. Rev. Lett.*, 100:070502, Feb 2008, arXiv:0704.3906.

- [80] D. Yang. A simple proof of monogamy of entanglement. *Physics Letters A*, 360(2):249–250, 2006, [arXiv:quant-ph/0604168](#).
- [81] D. Yang, K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, and W. Song. Squashed entanglement for multipartite states and entanglement measures based on the mixed convex roof. *IEEE Trans. Inf. Th.*, 55(7):3375–3387, July 2009, [arXiv:0704.2236](#).
- [82] K. Yasuda and H. Nakatsuji. Direct determination of the quantum-mechanical density matrix using the density equation. II. *Phys. Rev. A*, 56:2648–2657, Oct 1997.